

# **Projet de mise en place d'un circuit de traitement et de diffusion des thèses électroniques au Service Commun de la Documentation de Paris III : aspects techniques**

**Gestion de projet**

**Laurence Sunyach**

Correspondant d'établissement  
Dominique Minquilan  
Conservateur,  
SCD Université Paris III Sorbonne nouvelle

Sous la direction de Catherine Cyrot  
Professeur associé à l'Enssib

## ***Remerciements***

Je remercie tout particulièrement Madame H  l  ne Chaudoreille, directrice du SCD de Paris III, qui m'a sugg  r   ce sujet, Madame Dominique Minquilan pour son suivi attentif, Mademoiselle Ga  la Bru pour sa relecture, ainsi que les nombreuses personnes qui m'ont re  ue, aid  e et orient  e dans la pr  paration de ce projet.

Mes remerciements s'adressent   galement    Madame Catherine Cyrot pour ses pr  cieux conseils ainsi qu'aux   l  ves conservateurs de l'ENSSIB, Anne-Claire Bernaudin, Camille D  gez , Lucile Dupuich, Nicole Feuchot, Claire Nguyen, Nicolas Pinet pour leur   tude tr  s document  e de la mise en place du circuit dans ses aspects administratifs et juridiques.

## **Résumé :**

Aujourd'hui, la valorisation de la recherche est un enjeu important pour l'Université française. Dans ce cadre, le SCD de l'université Paris III souhaite mettre en place une chaîne de traitement et de mise en ligne des thèses. Ce projet étudie les aspects techniques de la mise en œuvre de ce nouveau circuit, à partir des fonctions traditionnelles d'une bibliothèque : collecter, signaler, diffuser et conserver des documents.

## Descripteurs :

Thèses et écrits académiques – Numérisation -- France

Université de la Sorbonne nouvelle (Paris)

Toute reproduction sans accord express de l'auteur à des fins autres que strictement personnelles est prohibée.

**Abstract :**

Improving visibility and potential exploitation of data produced by basic science represents an important goal for French University. The Academic Library of the Paris III University aims at establishing digitization and electronic publication of theses and dissertations. The present report examines technical aspects and set up of such facilities starting from traditional functions of a library ie: collecting, cataloging, and conservation.

Keywords :

Dissertations – Digitization – France

# Sommaire

INTRODUCTION .....	9
<b>PARTIE 1 UN PROJET POUR UN ETABLISSEMENT .....</b>	<b>10</b>
1. LE CONTEXTE : L'UNIVERSITÉ PARIS III ET LA DOCUMENTATION : VERS DE NOUVEAUX PARTENARIATS .....	10
1.1. <i>Paris III Sorbonne nouvelle : une université en mutation</i> .....	10
1.1.1. Histoire et statuts.....	10
1.1.2. Organisation et composantes : une université éclatée.....	11
1.1.3. La recherche à l'université de Paris III .....	11
1.2. <i>Le service commun de la documentation</i> .....	12
1.2.1. Les statuts et les missions.....	12
1.2.2. La répartition géographique, les ressources et le personnel.....	13
1.2.3. Les collections .....	14
2. LES ENJEUX NATIONAUX ET LOCAUX .....	14
2.1. <i>Mieux valoriser la recherche universitaire française</i> .....	14
2.1.1. Les thèses premier vecteur de la recherche .....	14
2.1.2. Un potentiel documentaire à valoriser en améliorant le signalement et la diffusion .....	15
2.1.3. Le dispositif réglementaire national : une forte incitation, de nombreuses conséquences.....	16
2.2. <i>La mise en ligne des thèses : un défi pour l'université et son SCD</i> .....	19
2.2.1. Les axes de développement du SCD de Paris III : mettre en place des services novateurs.....	19
2.2.1.1. Améliorer les services aux lecteurs .....	19
2.2.1.2. Collaborer et renforcer la visibilité au sein de l'université .....	20
2.2.2. Genèse et enjeux du projet à l'université Paris III : besoins et atouts .....	21
2.2.2.1. Identification des besoins.....	21
Volumétrie et spécificité du fonds .....	21
Faciliter la consultation et mieux conserver .....	22
2.2.2.2. Les atouts et les acteurs du projet de l'université Paris III.....	23

Les acteurs et les partenaires potentiels .....	23
La méthodologie : une étude à deux volets.....	24
3. LES EXEMPLES EXTÉRIEURS.....	25
3.1. <i>A l'échelle internationale</i> .....	25
3.1.1. Virginia Tech et le NDLTD : un temps d'avance .....	25
3.1.2. E-theses la mise en ligne des thèses à l'université d'Utrecht : un exemple européen .....	26
3.2. <i>En France</i> .....	28
3.2.1. Quelques exemples de production en XML.....	28
3.2.1.1. Cyberdocs : la chaîne de production de l'université Lumière Lyon 2 .....	28
3.2.1.2. SPARTE : l'ABES .....	29
3.2.1.3. ... et quelques autres : INJAC et CASTORE .....	30
3.2.2. Un outil de gestion pour l'université Lumière Lyon 2 : OGET ...	31

## **PARTIE 2 DU DEPOT A LA DIFFUSION : NOUVELLES**

### **PROBLEMATIQUES BIBLIOTHECONOMIQUES .....33**

1. LE PROJET DE MISE EN LIGNE DES THÈSES À PARIS III.....	33
1.1. <i>Repenser le circuit des thèses</i> .....	33
1.1.1. Le circuit actuel.....	33
1.1.2. Les objectifs à atteindre.....	34
1.2. <i>Les formats en question</i> .....	35
1.2.1. Le format PDF .....	35
1.2.2. Le langage XML .....	36
1.2.3. Le choix de Paris III.....	37
1.3. <i>La nouvelle organisation</i> .....	38
1.3.1. UniTEL : la collaboration des services universitaires .....	38
1.3.2. Les moyens et les limites.....	39
1.3.2.1. Les outils informatiques.....	39
1.3.2.2. Le personnel disponible .....	39
1.3.3. Les données variables non connues .....	40
2. LE TRAITEMENT DOCUMENTAIRE : GESTION DU DOCUMENT, SIGNALEMENT, ACCÈS, DIFFUSION ET CONSERVATION .....	40

2.1.	<i>La modélisation du workflow</i> .....	41
2.2.	<i>Le signalement</i> .....	42
2.2.1.	Interopérabilité : le protocole OAI.....	42
2.2.2.	Les catalogues et les bases de données .....	43
2.2.3.	Les accès aux bases .....	44
2.3.	<i>La diffusion</i> .....	44
2.3.1.	Modalités de la diffusion .....	44
2.3.2.	Le format de diffusion .....	45
2.4.	<i>La conservation</i> .....	46
3.	LES OUTILS DU PROJET .....	46
3.1.	<i>Le choix d'un outil de conversion</i> .....	47
3.2.	<i>Le signalement, la diffusion et la conservation</i> .....	48
3.3.	<i>Les questions en suspens: conséquences de la mise en production de STAR</i> .....	49
	<b>PARTIE 3 LA MISE EN ŒUVRE .....</b>	<b>52</b>
1.	MISE EN PLACE DE LA CHAÎNE DE TRAITEMENT .....	52
1.1.	<i>Le nouveau circuit des thèses</i> .....	52
1.1.1.	Du dépôt à la mise en ligne .....	52
1.1.2.	Schéma général du circuit.....	54
1.2.	<i>Des métiers, des acteurs : répartition des tâches et implications des différents services</i> .....	55
1.2.1.	Le doctorant .....	55
1.2.2.	Les écoles doctorales.....	55
1.2.3.	Le service des doctorats .....	55
1.2.4.	Le jury .....	55
1.2.5.	Le service commun de la documentation .....	55
1.2.6.	Le centre audiovisuel et informatique (CAVI).....	56
1.2.7.	Le centre de ressources informatiques (CRI) .....	56
1.2.8.	Autres intervenants extérieurs .....	56
1.2.9.	Articulation des applications : une hypothèse basse sans application de workflow.....	56
2.	RÉALISATION DU PROJET : PROGRAMMATION .....	57

2.1.	<i>Organisation des tâches au sein du SCD et fiches de postes</i>	57
2.1.1.	Le chef de projet	57
2.1.2.	Le Technicien	58
2.1.3.	Le styleur	58
2.2.	<i>Le matériel</i>	59
3.	LE CALENDRIER ET LE SUIVI	60
3.1.	<i>Le «phasage»</i>	60
3.2.	<i>Le calendrier</i>	61
3.3.	<i>Evaluation et suivi</i>	63
3.4.	<i>Perspectives et avenir</i>	64
3.4.1.	Au sein de l'université :	64
3.4.2.	Vers un pôle universitaire	65
	<b>CONCLUSION</b>	<b>66</b>
	<b>BIBLIOGRAPHIE</b>	<b>67</b>
	<b>TABLE DES ANNEXES</b>	<b>71</b>

## ***Introduction***

Premier vecteur des résultats de la recherche, les thèses françaises sont aujourd'hui mal valorisées. Avec Internet, les bibliothèques électroniques se constituent. Grâce aux protocoles d'échanges de données qui rendent les archives électroniques visibles du monde entier, la diffusion électronique des thèses est devenue un mode de valorisation efficace.

La circulaire ministérielle parue au printemps 2005 a donnée une formidable impulsion à un mouvement débuté en 2000 par quelques établissements pionniers. Les projets renaissent et se multiplient. Déjà très engagé dans la documentation électronique, le SCD de Paris III a souhaité se lancer dans la diffusion électronique des thèses. Je désirais vivement un sujet lié à ce nouveau champ du métier de bibliothécaire, j'ai donc trouvé dans ce projet passionnant un prétexte au renouvellement de mes connaissances.

Les aspects administratifs et juridiques du circuit de mise en ligne des thèses ont fait l'objet d'une étude menée cette année par un groupe projet du DCB 14 de l'ENSSIB. Mon travail s'appuie donc sur les propositions qui en résultent. Il aborde les aspects techniques de la mise en ligne sous l'angle des problématiques bibliothéconomiques.

Après avoir présenté le contexte et les enjeux du projet, nous nous servons des expériences existantes pour étudier sa faisabilité et les outils possibles au regard des objectifs du SCD. Puis dans une troisième partie, nous établirons des propositions pour sa mise en œuvre.

# ***Partie 1 UN PROJET POUR UN ETABLISSEMENT***

## **1. Le contexte : l'université paris III et la documentation : vers de nouveaux partenariats**

### **1.1. Paris III Sorbonne nouvelle : une université en mutation**

#### **1.1.1. Histoire et statuts**

Fondée en 1970, Paris III est une université récente. Sa création fait partie du programme conçu après les événements de mai 1968, pour remédier à l'extrême précarité de l'enseignement supérieur français en général, et parisien en particulier. Son projet repose sur la volonté d'ouvrir l'institution universitaire aux mutations de la société de l'époque, en intégrant les nouvelles approches disciplinaires de la recherche dans l'enseignement des arts, des langues, des lettres et des sciences humaines. Très influencée par le mouvement de modernisation de l'Université, elle reste néanmoins attachée à la tradition académique de la Sorbonne. En 1971, elle prend le nom de « Sorbonne nouvelle » pour marquer cette double filiation et son implantation dans le Quartier Latin.

L'université de la Sorbonne nouvelle est un établissement public à caractère scientifique culturel et professionnel, EPSCP dont le fonctionnement est régi par la loi du 26 janvier 1984 relative à l'enseignement supérieur. Elle est dirigée par un président élu pour un mandat de cinq ans par l'ensemble des membres des trois conseils : conseil d'administration, conseil scientifique, conseil des études et de la vie universitaire.

Le président est assisté d'un bureau, des 3 vice-présidents des conseils, de la vice-présidence déléguée aux relations internationales, de chargés de missions, du secrétaire général et de l'agent comptable. Monsieur Bernard Bosredon, professeur

de sciences du langage, est Président de l'université Paris III depuis le 14 octobre 2002.

### 1.1.2. Organisation et composantes : une université éclatée

L'université accueille chaque année environ 20 000 étudiants de toutes nationalités, inscrits dans les différents cursus. Les effectifs sont constants d'une année sur l'autre. En 2004-2005, ils étaient 7347 en premier cycle, 8026 en deuxième cycle (licence, maîtrise, et préparation au capes et à l'agrégation) et 3258 en troisième cycle. L'offre de formation recouvre 110 diplômes nationaux et 20 diplômes d'université<sup>1</sup>. 544 enseignants-chercheurs et enseignants dispensent les cours et encadrent les recherches.

L'université Paris III se caractérise par son éclatement géographique sur 10 sites répartis entre Paris et sa proche banlieue. Néanmoins, la majorité des unités de formation (UFR) et des services sont localisés dans le Quartier Latin et à proximité, sur le site Censier. L'ensemble de ces implantations couvre une surface totale de 41 008 m<sup>2</sup> qui ne permet pas d'offrir des conditions de travail optimales. Des travaux de désamiantage des bâtiments sont indispensables. Enfin, le rapprochement des différentes composantes rendrait le fonctionnement de l'université plus rationnel. Les projets de déménagements se succèdent mais l'avenir reste incertain.

### 1.1.3. La recherche à l'université de Paris III

Le contrat quadriennal 2005-2008 verra la rénovation des dispositifs et des contenus de formation. A partir de septembre 2005, les étudiants poursuivront leurs études dans le cadre du LMD inscrit dans un seul domaine de formation intitulé : « Arts, langues, lettres, sociétés contemporaines ».

La recherche est représentée au sein de 51 unités de recherche et de 7 écoles doctorales :

- (Écoles doctorale 120) Littérature française et comparée
- (Écoles doctorale 122) Europe latine – Amérique latine

- (Écoles doctorale 265) Langues et civilisations orientales
- (Écoles doctorale 267) Arts du spectacle, sciences de l'information et de la communication
- (Écoles doctorale 268) Langage et langues : description, théorisation, transmission
- (Écoles doctorale 384) Etudes anglophones
- (Écoles doctorale 385) Espace européen contemporain : politiques, économies, société, culture

Pour structurer la recherche et renforcer le rôle des Écoles doctorales, Paris III soutient le projet de création d'une maison de la recherche assistée d'un collège des Écoles doctorales.

## **1.2. Le service commun de la documentation**

### **1.2.1. Les statuts et les missions**

Le service commun de la documentation a été créé conformément au décret de 1992 pour servir la communauté des étudiants, enseignants et chercheurs de l'université Paris III. Ses collections sont développées en fonction de l'évolution des enseignements et de la recherche en lettres, langues, sciences humaines, théâtre et cinéma. Elles coexistent avec les fonds, plus encyclopédiques, des bibliothèques Sainte Geneviève, Sainte-Barbe, des Langues Orientales, BIULO, toutes trois placées sous l'autorité administrative de l'université Paris III. Les collections de la BIULO ainsi que celles de la bibliothèque de l'Institut des études iraniennes et turques rejoindront la future bibliothèque universitaire des langues et civilisation actuellement en préfiguration.

Conformément au Décret n° 85-694 du 4 juillet 1985, le SCD a pour mission de :

- mettre en œuvre la politique documentaire de l'Université, coordonner les moyens correspondants et évaluer les services offerts aux usagers.
- acquérir, gérer et communiquer les documents de toute nature qui appartiennent à l'Université ou qui sont à sa disposition.

---

<sup>1</sup> Source : Guide de l'étudiant 2004-2005 / Université Paris III Sorbonne Nouvelle

- participer, à l'intention des utilisateurs, à la recherche documentaire, à la production de l'information à sa diffusion ainsi qu'aux activités d'animation.
- favoriser par l'action documentaire et l'adaptation des services, toute initiative dans le domaine de la formation initiale et continue et de la recherche.
- coopérer avec les bibliothèques qui concourent aux mêmes objectifs, quels que soient leurs statuts, notamment par la participation à des catalogues collectifs en particulier dans le cadre d'un réseau régional.
- donner aux utilisateurs l'accès aux techniques nouvelles d'information scientifique et technique.

### 1.2.2. La répartition géographique, les ressources et le personnel

A l'instar de l'université, le SCD regroupe 10 unités documentaires réparties sur différents sites. La place centrale de ce réseau est occupée par la bibliothèque universitaire dont une section, la bibliothèque d'allemand, est distante (Annexe 1). Sept bibliothèques d'UFR ont été progressivement intégrées : la bibliothèque du monde anglophone, la bibliothèque des études italiennes et roumaines, la bibliothèque de linguistique et phonétique générale et appliquée, la bibliothèque de linguistique africaine, la bibliothèque des études portugaises, brésiliennes et d'Afrique lusophone, la bibliothèque de littérature générale et comparées. Les trois autres : la bibliothèque des études théâtrales Gaston Baty, l'Institut des hautes études d'Amérique latine, la bibliothèque des études iraniennes et turques sont associées. Toutes participent à l'élaboration du catalogue collectif.

Ce réseau documentaire mobilise un effectif total de 72 agents titulaires dont 38 sont affectées à la bibliothèque centrale, 16 dans les bibliothèques intégrées et 18 dans les bibliothèques associées. Pour assurer complètement les missions qui lui sont confiées, le SCD recrute chaque année du personnel vacataire ainsi que des moniteurs étudiants.

La bibliothèque centrale compte 6 conservateurs, 2 bibliothécaires, 9 bibliothécaires adjoints et assistants de bibliothèques, 1 SASU, 15 magasiniers en chef et magasiniers spécialisés, 5 agents et adjoints techniques.

Un adjoint technique de la branche d'activité professionnelle E, personnel du CRI mais installé dans les locaux du SCD, est chargé de la maintenance du parc de micro-ordinateurs du SCD (140 postes répartis entre les postes professionnels, postes publics, postes de formation). Il est aidé par un assistant de bibliothèque. Tous deux travaillent sous la responsabilité d'un conservateur chargé de l'administration des différents systèmes informatiques.

### 1.2.3. Les collections

Les collections comptent environ 280 000 ouvrages dont 50 % sont conservés à la bibliothèque universitaire. Le SCD est abonné à 1080 titres de périodiques. Depuis 2001, l'offre documentaire intègre la documentation électronique. Le SCD de Paris III a ainsi été parmi les premières bibliothèques de lettres et sciences humaines à proposer une offre en ligne significative avec environ 12000 titres de périodiques et 130 ouvrages de références. La documentation électronique est un axe fort de la politique documentaire (environ 30% du budget).

## 2. Les enjeux nationaux et locaux

### 2.1. Mieux valoriser la recherche universitaire française

#### 2.1.1. Les thèses premier vecteur de la recherche

Les thèses sont actuellement un vecteur fondamental de la communication de la recherche universitaire. Rédigées au sein d'une École doctorale, selon les indications d'un directeur, elles sont le fruit d'une collaboration. Elles reflètent le dynamisme d'une équipe de recherche ou d'un laboratoire et contribuent ainsi à la renommée d'une université. Aujourd'hui, pour publier leur thèse, les jeunes auteurs disposent de peu de canaux : les Presses universitaires, le circuit

traditionnel de l'édition et dans certains cas la diffusion électronique. Lorsqu'il n'existe aucun dispositif de mise en ligne au niveau institutionnel, université ou établissement habilité, ils ont recours à des systèmes alternatifs comme par exemple un site personnel, ou le Centre pour la communication scientifique directe, CCSD<sup>2</sup>.

Les chercheurs qui travaillent dans les disciplines littéraires ou les sciences sociales ont encore moins que les autres l'occasion de valoriser leurs travaux. L'université doit s'engager pour la valorisation des thèses : c'est un enjeu fondamental pour la vitalité de la recherche française en sciences humaines et sociales.

### 2.1.2. Un potentiel documentaire à valoriser en améliorant le signalement et la diffusion

En novembre 1997, *l'Enquête sur les pratiques des utilisateurs du signalement des thèses et des utilisateurs des thèses* confiée par le Ministère de l'Education Nationale à la société SCP Communication, montrait l'insatisfaction des usagers. Le rapport indique que les lecteurs « souhaitent manifestement avoir accès rapidement et facilement à l'information secondaire (bibliographique) correspondante, pouvoir apprécier dans les mêmes conditions si une thèse repérée est de nature à leur être utile dans leur travail, pouvoir consulter le texte intégral et, le cas échéant, pouvoir passer commande d'un substitut (photocopies, microformes, éditions à la carte) ».

Participer à la diffusion des résultats de la recherche de la communauté scientifique est une des missions traditionnelle d'un SCD. Le mode de diffusion actuel des travaux de recherche mis en place par le Ministère de l'Education nationale en 1985 valorise très peu les thèses. Elles sont signalées principalement par deux types d'outils bibliothéconomiques : les catalogues, dont le catalogue collectif des thèses françaises, Docthèses<sup>3</sup>, est intégré au système universitaire de documentation SUDOC depuis son déploiement en 2000, et les bases de données spécialisées.

---

<sup>2</sup> Le CCSD (CNRS), développe une archive électronique qui comprend une collection de thèses basée sur le serveur TEL. Les doctorants peuvent procéder à un dépôt volontaire de leur travail. Les documents sont diffusés après un contrôle rapide de la validité de l'information. Cette diffusion n'a aucun caractère officiel. Les universités les plus représentées dans la base sont des universités scientifiques.

Pour la plupart, les thèses sont communiquées dans les bibliothèques qui dépendent des universités dans lesquelles elles ont été soutenues, et dans les bibliothèques des disciplines proches sous forme de microfiches. Leur consultation se fait sur place, pour des questions de droit et de conservation. Leur circulation reste limitée au PEB. Elles sont en général peu consultées : une fois tous les dix ans en moyenne.<sup>4</sup>

Avec le SUDOC, la qualité du signalement des thèses a progressé. Le passage du support cédérom au catalogue consultable sur Internet a permis d'améliorer à la fois sa mise à jour et sa diffusion. Les thèses sont très diversement représentées dans les bases de données qui couvrent inégalement les disciplines. Cela ne constitue pas une valorisation suffisante notamment pour la promotion de la recherche française à l'étranger.

### 2.1.3. Le dispositif réglementaire national : une forte incitation, de nombreuses conséquences

Le dépôt, le signalement, la diffusion des thèses en France est encadré par un ensemble de textes réglementaires dont l'application est confiée à la Direction de l'Enseignement Supérieur.

Le Ministère de l'Education nationale a, depuis le rapport Jolly, manifesté sa volonté de valoriser et de pérenniser les thèses françaises grâce à leur diffusion par voie électronique. Cette position officialisée dans la circulaire n° 2000-149 du 28 septembre 2000, vient d'être confirmée par la circulaire n° 05- 94 du 29 mars 2005 dans laquelle est annoncée l'obligation prochaine du dépôt électronique des thèses (III,2). Il ne s'agit pas d'une règle pour l'instant (ni loi, ni décret) mais plutôt de l'expression d'un projet en forme de forte incitation.

En 2000, la circulaire prenait en compte les aspects dépôt et diffusion des thèses. Elle recommandait :

- le dépôt électronique

---

<sup>3</sup> Catalogue de toutes les thèses soutenues en France depuis 1972 (1983 pour les disciplines de santé)

- la conversion du document au format XML permettant un archivage pérenne
- la diffusion en format HTML ou PDF
- le signalement dans le Sudoc
- l'archivage local et/ou au CINES
- la diffusion institutionnelle

A l'échelon national, quelques universités pionnières se sont lancées dans l'aventure. Un premier bilan de ces expériences a été dressé dans le rapport rendu à la Direction de l'Enseignement Supérieur, en 2004 par la société Six et dix. Les conclusions de ce rapport montrent un application limitée du dispositif : seulement 645 thèses étaient en ligne au 2<sup>e</sup> trimestre 2004 dont 360 en format XML. Quatorze établissements ont organisé un dépôt mixte, papier et électronique, huit utilisent une chaîne de conversion XML. Ces résultats mitigés sont en grande partie dus à des problèmes structurels qui ne permettent pas de faire face à un choix technique très lourd. Il y a peu de moyens humains, le temps de traitement lié à la conversion au format XML est très long (2 à 3 jours par thèse). Dans ces conditions, il faudrait 150 ETP (Equivalent temps plein) pour traiter les quelques 10 000 thèses produites chaque année dans les Universités et les établissements habilités à décerner le grade de docteur. Il manque du temps et des moyens pour résoudre les bugs de conversion des chaînes principales : Sparte et Cyberdocs.

A la suite de ce constat un nouveau dispositif de diffusion électronique des thèses par voie électronique a été élaboré. Ces préconisations ont été décrites dans une circulaire très attendue par les établissements déjà engagés dans la démarche ou souhaitant mettre leurs thèses en ligne.

En 2005, la circulaire qui prend acte des difficultés observées, assoupli le dispositif et préconise toujours :

- le dépôt électronique

---

<sup>4</sup> Chiffre extrait d'une communication de Jean Paul Ducasse présentée lors de la journée d'étude sur la diffusion électronique des thèses organisée le 7 juillet 2005 à Paris par l'ADBS.

- le signalement dans le SUDOC
- l'archivage au CINES
- la diffusion en priorité au sein de l'établissement

Mais la structuration en XML n'est plus décisive. En effet, la rapidité de diffusion devient un enjeu prioritaire. Le choix du modèle de conversion est possible mais à condition de prendre en compte le temps de traitement de la thèse. Les établissements maîtrisent leur circuit, toutefois le ministère introduit deux contraintes majeures par le biais de l'utilisation d'une application nationale de dépôt, dont la mise en production annoncée pour la fin de l'année 2005 est confiée à l'Agence Bibliographique de l'Enseignement Supérieur et l'utilisation de métadonnées normalisées en cours de définition à l'AFNOR. Cette démarche s'appuie sur la nature administrative de la thèse et cherche à renforcer la validation du document mis en ligne. Déposée et diffusée via l'application nationale, la thèse électronique devient le document officiel et canonique en lieu et place du document papier. On considère que le format PDF est plus sûr que le format XML pour résoudre les problèmes d'archivage et de conservation, aussi l'utilisation de l'application nationale et le stockage au CINES deviennent-elle presque inévitables pour qui choisira le format PDF.

Un nouvel arrêté tenant compte de ces nouveaux paramètres est en préparation. Il devrait venir remplacer celui du 25 septembre 1985 qui régit actuellement le circuit de dépôt des thèses. Toutefois, le ministère annonce une nouvelle circulaire à paraître à la fin de l'année 2005.

Ces orientations nationales ont un impact direct sur les universités françaises. Elles concernent en effet tous les acteurs impliqués dans la vie d'une thèse depuis sa création jusqu'à sa mise à la disposition du public. Elle suppose ainsi une réorganisation complète du circuit administratif et une redéfinition des interventions de chaque services.

## **2.2. La mise en ligne des thèses : un défi pour l'université et son SCD**

### **2.2.1. Les axes de développement du SCD de Paris III : mettre en place des services novateurs**

Le SCD de Paris III s'est doté d'une charte des collections, validée par le conseil de la documentation en novembre 2004<sup>5</sup>. Elle décrit les missions de la bibliothèque en matière de documentation et les principes qui président à la constitution de ses collections. Le SCD participe par ailleurs au plan de développement des collections du Quartier Latin impulsé par la Sous-Direction des Bibliothèques de l'Enseignement Supérieur avec pour objectif de rationaliser la répartition des collections entre les divers pôles documentaires du secteur.

#### *2.2.1.1. Améliorer les services aux lecteurs*

Comme l'université Paris III dont il fait partie, le SCD souffre de la pression du nombre des étudiants et du manque de surface. Cette contrainte oriente fortement sa politique générale.

Pour faire face à la saturation des espaces de stockage, l'équipe a engagé un travail à long terme sur les fonds : récolement, désherbage, déstockage. Ainsi, les ouvrages peu consultés sont stockés à Marne-la-Vallée au Centre technique de l'enseignement supérieur, CTLES et communiqués de façon différée. Les thèses sont des documents volumineux, souvent peu consultés, le premier envoi a donc concerné les plus anciennes d'entre elles.

Pour améliorer les conditions d'accueil du public et notamment le manque de postes informatiques en usage libre, la salle de lecture a été équipée d'un réseau WIFI qui permet aux usagers disposant d'un ordinateur portable d'utiliser leurs propres outils bureautiques.

La bibliothèque propose des services spécifiques aux étudiants en situation de handicap. Ainsi en 2004, elle a acquis du matériel informatique destiné aux personnes aveugles ou amblyopes (les plus représentées à l'université de Paris III).

---

<sup>5</sup> Source : Service commun de la documentation : rapport d'activité 2004

Grâce aux crédits accordés dans le cadre du prix Handinnov<sup>6</sup> 2003, un téléagrandisseur, un clavier azerty « cherry » agrandi, un logiciel zoomtext, des équipements pour non voyants : logiciel d'accès à windows NT : Jaws NT, logiciel de synthèse vocale SAY IT PRO claire, scanner Epson et son logiciel OCR OpenBook, et un terminal braille Alva Delphi ont été achetés. Une webcam permet aux personnes sourdes connaissant la langue des signes de « signer » avec une autre personne extérieure. Enfin, une table inclinable à hauteur variable a été installée dans la salle de lecture pour les personnes en fauteuil roulant.

Parallèlement aux services classiques<sup>7</sup>, la bibliothèque met en place une politique de services novateurs.

Le catalogue collectif baptisé MAESTRO (porté par l'application Aleph 500 de la société Ex Libris), accessible à l'adresse <http://maestro.scd.univ-paris3.fr> permet au lecteur d'effectuer des recherches parmi les fonds des 10 bibliothèques participantes et d'accéder à son compte personnel. Par le portail documentaire appelé VIRTUOSE, tous les utilisateurs inscrits, peuvent accéder en ligne et à distance aux ressources électroniques du SCD. Les lecteurs interrogent après identification, l'ensemble des ressources documentaires en plein texte dans une interface de recherche unifiée. VIRTUOSE (logiciel MetaLib de la société Ex Libris) est le premier portail documentaire disposant de la recherche fédérée et accessible à distance, installé dans une bibliothèque universitaire française.

#### 2.2.1.2. *Collaborer et renforcer la visibilité au sein de l'université*

En novembre 2004, le SCD et le conseil scientifique ont organisé une « exposition rencontre » avec les auteurs de l'université Paris III, leurs éditeurs et les médias. Cinq cents livres publiés entre 2000 et 2004 ont été présentés à cette occasion. Ils témoignent du dynamisme de la recherche à la Sorbonne nouvelle. Cette manifestation sera reconduite lors des assises de la recherche en mars 2006.

---

<sup>6</sup> Le concours «HANDINNOV» est organisé chaque année par L'ONISEP, DROIT AU SAVOIR et la revue RÉADAPTATION. Il distingue des réalisations et des pratiques innovantes ayant un impact sur l'insertion des jeunes personnes en situation de handicap dans leur vie scolaire, étudiante, professionnelle ou citoyenne au sens large.

<sup>7</sup> Prêt à domicile, prêt entre bibliothèques, photocopies, accès à Internet et aux bases de données, renseignements bibliographiques et formations à la recherche documentaire.

Depuis début 2005, le SCD expérimente une politique d'acquisition centralisée de toute la documentation de l'université comprenant les documents à l'usage des services internes.

Le SCD s'est doté au titre du contrat en cours, d'un outil de gestion des objets numériques (mise en ligne, description bibliographique, indexation-matière et mise en place de points d'accès dans un catalogue en ligne, gestion des droits des utilisateurs ...) dans laquelle les composantes de l'université pourront travailler pour mettre à la disposition des étudiants le résultat des programmes de numérisation engagés.

Le projet de diffusion électronique des thèses correspond parfaitement à cette forte volonté de développement des services à distance. Il permet aussi une meilleure visibilité du SCD dans la communauté universitaire.

### 2.2.2. Genèse et enjeux du projet à l'université Paris III : besoins et atouts

Le projet de l'université Paris III s'inscrit dans le cadre réglementaire exposé plus haut. Il en épouse à la fois les contraintes et les incertitudes. Mais il répond avant tout à des besoins qu'il convient maintenant de présenter.

#### 2.2.2.1. Identification des besoins

##### **Volumétrie et spécificité du fonds**

Le fonds des thèses de Paris III est évalué à 3500 titres. Il s'accroît d'environ 160 unités par an. Le volume n'a pas subi de variations notables depuis 2000.

	<b>ED 120</b>	<b>ED 122</b>	<b>ED 265</b>	<b>ED 267</b>	<b>ED 268</b>	<b>ED 384</b>	<b>ED 385</b>	<b>nb de thèses soutenues</b>
2000	56	28	8	33	34	22	7	<b>188</b>
2001	41	39	9	18	24	27	12	<b>170</b>
2002	41	13	9	20	27	15	8	<b>133</b>
2003	35	36	5	23	20	19	16	<b>154</b>

2004	41	30	9	23	31	17	12	<b>163</b>
------	----	----	---	----	----	----	----	------------

Evolution des thèses soutenues depuis 2000

Les sujets sont répartis entre les 7 écoles doctorales habilitées à suivre les recherches. La plupart des thèses comportent des tableaux, des figures, et de nombreuses illustrations (cartes et photographies). Certaines contiennent des caractères non latins. Quelques doctorants déposent des cédéroms ou des CD audio d'accompagnement qui comprennent des documents sonores ou des illustrations.

### **Faciliter la consultation et mieux conserver**

L'université Paris III est très attachée à la coopération internationale. Elle accueille de nombreux étudiants et enseignants chercheurs dans le cadre de programmes d'échanges : ainsi, les étudiants étrangers représentent 46% des effectifs des doctorants. Certaines des thèses soutenues à Paris III sont des thèses en cotutelle. On peut émettre l'hypothèse que ces étudiants seront désireux de voir leur travail plus amplement diffusé que dans le simple périmètre des universités françaises et dans une moindre mesure étrangères via le prêt entre bibliothèques.

Généralement, la conservation des collections en magasin ne favorise pas la consultation. Cela ne semble pas être le cas des thèses de Paris III qui représentent près d'un tiers des communications d'ouvrages du magasin. Les statistiques de consultations des thèses sur la période 2001-2004 sont en constante augmentation (Annexe 2). En 2000, pour libérer des mètres linéaires dans le magasin, les thèses dont la date de soutenance est inférieure ou égale à 1985 ont été envoyées au Centre technique de l'enseignement supérieur de Marne la Vallée, CTLES pour stockage. Une navette établie entre le CTLES et la bibliothèque permet de les communiquer à la demande en 2 ou 3 jours. La rotation de cette collection est loin d'être négligeable.

Quels sont les usages de ce corpus ? De façon très empirique, on constate que les demandes effectuées au bureau de renseignements sont parfois motivées par un besoin très formel : les étudiants souhaitent consulter une bibliographie, ou simplement savoir comment présenter leur travail. Cependant, ils les parcourent

aussi à partir de la table des matières à la recherche d'un concept ou d'une idée précise. Le support numérique permet, pour cet usage, une amélioration ergonomique considérable. Une récente étude menée par l'École des Mines de Nantes a montré l'importance des documents électroniques pour les chercheurs et les enseignants de cette école mais elle ne décrit pas le degré de finesse de l'utilisation du texte par les lecteurs.

Le SCD de Paris III développe depuis quelques années une politique très volontariste en direction des usagers souffrant de handicaps. Pour l'instant, les thèses constituent un fonds très difficile d'accès pour ce public. Pour un lecteur mal voyant la consultation d'une thèse ne peut se faire qu'en prenant rendez-vous, ou avec l'aide d'une personne extérieure. Une fois mises en ligne, les thèses pourront être consultées sur l'appareil à synthèse vocale ou le télé-agrandisseur, installés dans la salle de lecture. Ce corpus constitue une source d'information supplémentaire.

Les thèses françaises sont des collections fragiles. Leur édition est en général laissée aux soins des doctorants. Elles se présentent sous forme de volumes en format A4, reprographiés et thermocollés qui résistent mal au stockage et aux manipulations. Les thèses imprimées sont souvent reproduites voire volées. Elles sont l'objet de la plupart des infractions en matière de reprographie. Au cours de l'année 2005, suite à un trafic de cartes d'étudiants, des thèses de l'université ont ainsi été volées à deux reprises

Par ailleurs, contrairement aux craintes généralement exprimées, la diffusion électronique n'accroît pas les risques de plagiat. Au contraire, une thèse diffusée sur Internet sera plus rapidement identifiée. Un copié-collé dans un moteur permet d'attribuer facilement les idées. Ce changement de support peut donc constituer une mesure de protection.

#### 2.2.2.2. *Les atouts et les acteurs du projet de l'université Paris III*

### **Les acteurs et les partenaires potentiels**

Administrativement, le projet se place sous la tutelle de l'université. La décision de dépôt et de diffusion électronique des thèses relève de la politique de l'établissement, et revient donc au conseil d'administration. Mais elle concerne également le conseil des études, et de la vie universitaire (CEVU), sur qui repose l'organisation des études ainsi que le conseil scientifique (CS) qui doit être consulté pour toutes les questions relatives à la recherche.

Par ailleurs, on trouve dans les différents services des partenaires et des ressources, sur lesquels le projet peut s'appuyer. Les premiers services concernés sont le service des doctorats, interlocuteur administratif privilégié des doctorants tout au long de leur parcours, et les écoles doctorales, placées au cœur du dispositif pédagogique. Le centre audiovisuel et informatique (CAVI), qui dispose de salles de formations équipées et de formateurs constitue aussi un support possible. Le centre de ressources informatiques (CRI), responsable du réseau et de l'administration des systèmes informatiques centraux pourra fournir une expertise et éventuellement du matériel. Enfin le service de la reprographie et les Presses de la Sorbonne nouvelle font également figure de partenaires potentiels.

La réussite du projet dépend de la capacité du SCD à mobiliser l'ensemble de ces ressources.

### **La méthodologie : une étude à deux volets**

Au moment où paraissait au BOEN la circulaire n°2000-149, ni l'université, ni le SCD n'avaient souhaité réellement être porteur du projet. Le dispositif en effet paraissait lourd à mettre en place sans moyens financiers ou humains supplémentaires. La diffusion électronique des thèses à l'université de Paris III avait donc été reportée. En 2005, l'imminence de la parution de la nouvelle circulaire promettant une meilleure définition des rôles de chacun et l'opportunité de ma formation à l'ENSSIB ont fourni le prétexte à le relancer.

Déjà impliqué dans le traitement, la communication et la conservation des thèses sur support papier, le SCD apparaissait certes comme le porteur de projet le plus naturel. Cependant, sa mise en œuvre ne pouvait pas dépendre de lui seul. La diffusion électronique des thèses a de nombreuses conséquences administratives et techniques. Elle requiert également des compétences juridiques nouvelles. Tout cela ne pouvait pas être analysé dans le seul cadre de ma formation. L'équipe de

direction a donc confié les études préalables, d'une part à une équipe projet de conservateurs du DCB 14, et d'en faire, d'autre part, le sujet de mon projet professionnel personnel.

La partie administrative, qui impliquait la recherche de contacts avec les différents partenaires potentiels, étude des aspects juridiques spécifiques, a été confiée à six élèves conservateurs. Conformément à leur lettre de mission, ils ont établi un état des lieux des expériences existantes en lettres et sciences humaines, élaboré une note de synthèse sur le droit et la documentation électronique, proposé un plan d'organisation et un budget pour la mise en place du projet dans son ensemble et formalisé les différentes étapes de la chaîne depuis les formations jusqu'à la mise en ligne<sup>8</sup>.

Cette étude a duré six mois, ponctués par trois comités de pilotage organisés successivement en mars, mai et juin 2005. Un des trois scénarii présentés a été validé et développé. Mon travail s'inscrit dans le prolongement du projet ENSSIB et s'appuie sur les propositions qui ont été faites. Il étudie les modalités techniques de la diffusion électronique des thèses à partir des compétences premières d'un SCD : le traitement documentaire.

### **3. Les exemples extérieurs**

#### **3.1. A l'échelle internationale**

##### **3.1.1. Virginia Tech et le NDLTD : un temps d'avance**

La réflexion sur la mise en place d'un service national de thèses en ligne a débuté aux Etats-Unis en 1987. Mais c'est grâce à l'Université Virginia Tech qui finança le développement de la première DTD<sup>9</sup> spécifiquement conçue pour décrire les thèses que le mouvement pris véritablement son essor. Depuis janvier 1997, tous les mémoires de maîtrise et doctorat soutenus à l'université Virginia Tech sont

---

<sup>8</sup> Anne-Claire Bernaudin, Camille Dégez, Lucile Dupuich, Nicole Feuchot, Claire Nguyen, Nicolas Pinet. Projet de mise en place d'un circuit de traitement et de diffusion des thèses électroniques au Service Commun de la Documentation de Paris III : aspect administratif et juridique. Villeurbanne : Enssib, 2005.

<sup>9</sup> Document Type definition

déposés sous forme de fichiers PDF ou SGML balisé selon la Document Type Definition Electronic electronic thesis and Dissertations, DTD ETD.

L'université Virginia Tech est à l'origine du Networked Digital Library of Theses and Dissertations, NDTLD. Le réseau composé majoritairement d'universités compte aujourd'hui 174 membres<sup>10</sup>. Le seul organisme français représenté à ce jour est le CCSD.

### 3.1.2. E-theses la mise en ligne des thèses à l'université d'Utrecht : un exemple européen

Au cours de sa formation à l'Enssib, un élève bibliothécaire d'Etat doit effectuer au moins un stage d'une durée d'un mois dans une structure documentaire extérieure. Pour mettre mon étude en perspective, j'ai choisi d'effectuer ce stage dans une bibliothèque néerlandaise : la bibliothèque d'Utrecht, *Universiteitsbibliotheek Utrecht (UBU)*, au sein du service de publication, Igitur publication and archiving service.

Une des missions de ce service est d'alimenter et de maintenir l'archive institutionnelle ouverte : *Digital scientific repository of Utrecht University*. La collecte et l'archivage des documents est assurée par la plate-forme logiciel DSpace, outil de Content Management « open source » développé conjointement par les bibliothèques du MIT, Massachusetts Institut of Technology et la société Hewlett-Packard (HP). L'archive est organisée en « communautés », à l'intérieur desquelles sont agrégés des groupes logiques de documents organisés en collections. Si certaines de ces collections font l'objet de dépôts ouverts, validés par le service, c'est le personnel d'Igitur qui a la charge du dépôt, de la soumission dans DSpace et de la mise en ligne des thèses.

Le dépôt électronique des thèses n'est pas encore obligatoire. L'université a donc choisi de soutenir la diffusion électronique en proposant un dédommagement financier. Environ 50 % des thèses soutenues sont actuellement déposées sous forme électronique.

---

<sup>10</sup> Source : [www.ndtld.org](http://www.ndtld.org)

L'édition de l'exemplaire de soutenance est assurée par les services de l'université à partir d'un fichier PDF. C'est ce même fichier qui est déposé à la bibliothèque. Il sert ensuite à la mise en ligne. La procédure de dépôt débute comme pour tous les documents au service des «entrées». Les fichiers informatiques, les formulaires et les autorisations de diffusion nécessaires sont enregistrés dans une base de données sous Access qui permet de suivre le traitement du document étape par étape. Les fichiers PDF, Postscript, HTML/XML, Ms-Word sont acceptés ; les fichiers Tex ou Latex, les thèses composées en Mac ou Unix, sont refusés ou convertis en PDF. La publication électronique prend la forme d'une page HTML qui donne accès au fichier du texte intégral de la thèse en format PDF découpé selon la table des matières pour faciliter la consultation.

Par soucis d'efficacité à court terme, le format PDF a été préféré au format XML. Néanmoins, conscients du peu de garanties offertes par le format PDF en matière de conservation, les promoteurs du projet ont imaginé des solutions alternatives. Ainsi le fichier intégral converti au format .txt accompagné de commentaires sur la conversion est déposé en même temps que le reste du matériel informatique dans l'archive<sup>11</sup>. Sa présence a pour but de garantir l'intégrité des données à long terme. La bibliothèque nationale des Pays-Bas a la charge d'émuler les applications nécessaires à la lecture future des fichiers informatiques. Pour garantir l'accès au texte, le système attribue une sorte d'URL universelle ou *Persistent identifier* (*handle System* pour DSpace).

Les inconvénients du projet résident assurément dans les accès et la diffusion. DSpace repose sur des normes internationales telles que Dublin Core ou le protocole OAI-PMH qui assurent l'interopérabilité et une bonne visibilité aux documents. Mais le signalement bibliographique est insuffisant. Les thèses électroniques ne sont pas cataloguées dans le système local (ALEPH 500), ni dans le catalogue collectif néerlandais. Les notices de signalement saisies dans DSpace réduisent les possibilités de recherche aux quinze éléments de Dublin Core. La navigation à l'intérieur du document se limite au feuilletage des chapitres.

Les solutions adoptées pour la diffusion électronique des thèses à l'université d'Utrecht sont représentatives des choix techniques effectués au Pays-Bas et dans de nombreux projets internationaux. Elles ne correspondent pas totalement aux objectifs de Paris III que nous exposerons plus loin et ne peuvent pas être appliquées. Elles confirment néanmoins l'importance du signalement et le soin qui doit être porté à la définition des métadonnées. Cette expérience constitue donc un élément de comparaison intéressant.

## 3.2. En France

### 3.2.1. Quelques exemples de production en XML

Le projet de l'université de Paris III, repose comme nous le verrons plus loin, sur le format XML. Cette étude exclue donc toutes les chaînes qui ne prévoient pas de conversion dans ce format. Les dispositifs de publications pourtant intéressants développés par Doc'INSA à Lyon (CITHER), ou les grandes écoles parisiennes, (PASTEL) ne seront pas présentés ici.

Dans les projets qui nous intéressent, le principe de la production du document « repose sur la définition d'un modèle qui identifie tous les éléments de données et qui décrit et structure l'ensemble du document selon une formalisation normalisée.(feuille de style d'un logiciel de traitement de texte) »<sup>12</sup> Le document est ensuite reformaté en XML selon un modèle de description normalisé DTD qui lui assurera une interdépendance par rapport à la plate-forme de lecture. C'est la condition nécessaire pour qu'il soit lisible à l'avenir. Pour l'instant, la mise au format XML passe par une chaîne de conversion.

#### 3.2.1.1. *Cyberdocs : la chaîne de production de l'université Lumière Lyon 2*

La plate-forme Cyberdocs est née de la collaboration entre les universités de Lyon 2 et de Montréal. Elle constitue l'outil de conversion d'un projet né en 1999 et baptisé Cyberthèses. Cette chaîne de production repose sur la norme SGML et la DTD TEI Lite. Depuis 2002, la chaîne est construite sur des outils XML,

---

<sup>11</sup> DSpace est conforme au modèle d'archivage OAIS, Open Archival information system

<sup>12</sup> Christian Lupovici, *La chaîne de traitement des documents numériques : caractéristiques et mise en œuvre*. in BBF, t. 47, n°1

entièrement en Open Source et sous licence GPL. Tous les établissements qui le souhaitent peuvent donc télécharger ses sources et installer la plate-forme qui peut être adaptée à leurs besoins. Utilisée principalement pour la diffusion des thèses, Cyberdocs peut traiter différents types de documents. Elle est documentée sur le site du CRU<sup>13</sup>. Les technologies utilisées s'appuient sur l'environnement informatique JAVA, XSLT, XSL-FO et OpenOffice.

La plate-forme Cyberdocs est constituée de trois modules. La transformation des documents de traitement de texte en format XML se fait dans le module de conversion. Le fichier Word ou OpenOffice est d'abord converti en un fichier « XML à plat ». Ce qui signifie qu'il ne présente aucune hiérarchie. Puis tous les éléments qui doivent être traités sont identifiés et transformés en éléments TEI Lite, en fonction des noms et des styles. La structure logique est reconstruite et le document XML final est produit. La plate-forme comporte en plus un module de publication qui pilote la diffusion et la recherche dans les documents, basé sur le moteur de recherche SDX et un module de gestion qui pilote la conversion et la publication des documents.

Remarque : en 1997, l'université de Marne-la-Vallée a mis en place, avec la société Jouvès un projet d'édition et de gestion des documents électroniques. Le module de conversion XML qui s'appuie sur la chaîne Cyberdocs est actuellement en cours de restructuration.

### 3.2.1.2. *SPARTE : l'ABES*

SPARTE, système de production et d'archivage de thèses électroniques, est la chaîne de traitement des thèses gérée par l'ABES. Conçue à partir des programmes sources de la chaîne cyberthèses avant son évolution, elle fonctionne sur les mêmes principes que celle-ci. Les développements ont été confiés à la société STILO ; éditrice d'OmniMark<sup>14</sup>. Elle a été testée par de nombreux établissements. Elle fonctionne de façon centralisée sur le modèle client serveur. Le fichier XML plat est produit à partir d'un document en format RTF. Le document est converti ensuite en XML TEI Lite puis en HTML pour la diffusion et le signalement via

---

<sup>13</sup> <http://sourcessup.cru.fr/Cyberthèses>

l'extracteur de données de WinIBW<sup>15</sup>. C'est un outil distant dont l'administration a lieu à l'ABES à Montpellier.

### 3.2.1.3. ... et quelques autres : INJAC et CASTORE

D'autres plate-formes existent. Parmi elles, on s'attachera principalement aux outils ouverts et à ceux destinés à la constitution d'un réseau.

#### INJAC :

INJAC est une plate-forme de gestion de contenus développée dans le cadre du projet ESUP-Portail<sup>16</sup>. C'est une chaîne de dépôt et de publication de littérature grise. La plate-forme utilise le système de publication XML du projet Apache Cocoon. La restitution des documents peut ainsi être faite dans différents formats : HTML, WML, PDF SVG, RTF. Elle doit être testée par le service commun de documentation de l'université Henri Poincaré Nancy 1 à partir du mois de septembre 2005. Mais cette expérimentation ne concerne pas la conversion au format XML. Cette fonctionnalité testée aussi par l'INSA de Lyon ne semble pour l'instant pas satisfaisante.

#### CASTORE :

CASTORE, Capitalisation, STockage et Recherche, est une plate-forme de bibliothèque numérique développée depuis 2001 par le centre de développement pédagogique de l'École des Mines de Nantes, l'IRISIA (INRIA Rennes) et Xerox Global Services. L'ensemble de la chaîne configurable repose sur les technologies J2EE/Java et composants Open source (Tomcat, Lucène...). Pour permettre une diffusion en format HTML ou PDF, la plate-forme convertit les documents et les métadonnées en format XML en utilisant la DTD Docbook. Déclarée Open Source depuis janvier 2005, elle est actuellement testée par la bibliothèque d'Angers. Le développement est désormais assuré par l'École des Mines de Nantes et le SCD de Nantes. Une des évolutions en cours concerne le passage à la TEI Lite. On dispose

---

<sup>14</sup> <http://www.stilo.com/>

<sup>15</sup> L'extracteur de données ne fonctionne plus dans la dernière version de WinIBW (3.0.72.2) mise à la disposition des établissements.

<sup>16</sup> ESUP-Portail est un environnement de travail numérique permettant l'accès à différents services, sources d'informations et ressources numériques.

assez facilement d'une plate-forme de test pour CASTORE. Le téléchargement de la plate-forme est possible sur le site : <http://www.emn.fr/x-info/bn/tele.htm>. On obtient un identifiant de test auprès de l'équipe de développement à la même adresse.

### 3.2.2. Un outil de gestion pour l'université Lumière Lyon 2 : OGET

Outre la conversion du document en format XML, il est nécessaire de disposer d'une application de dépôt et de suivi des documents au travers du circuit de traitement (workflow). Pour gérer ce workflow, on peut envisager d'utiliser les dispositifs existants ou de développer un module en interne.

L'université technologique de Compiègne utilise un module de dépôt développé par la société Ever, directement consultable sur son portail. D'autres utilisent des dispositifs souples tels qu'un fichier Excel ou à une base de données Access. Le service des publications électroniques de l'université Lyon 2 a développé un Outil de gestion électronique des thèses, OGET, qui inclue le workflow et sert de module de dépôt. Il est conçu pour préparer et suivre le traitement du document à travers la chaîne de conversion Cyberthèses. C'est une base PhpMySQL installée sur un serveur central interne. OGET fonctionne avec un accès sécurisé par login et un mot de passe sur les postes clients. Il est doté d'un moteur de recherche qui permet d'accéder à l'espace d'une thèse ou d'afficher une liste de thèses selon différents critères. C'est un outil libre qui est mis gratuitement à la disposition des établissements qui souhaitent l'utiliser.

Toutes les étapes du dépôt sont consignées dans OGET : validité du dépôt, des supports informatiques, présence des sources nécessaires à la bonne lisibilité du fichier, conformité de la version électronique avec le support papier, présence des autorisations et des formulaires nécessaires... Les fichiers stylés sont exportés vers la chaîne de conversion Cyberdocs. Les métadonnées nécessaires au signalement sont saisies, importées d'APOGEE<sup>17</sup> (système utilisé par la scolarité) ou extraites de documents officiels (procès verbal de soutenance par exemple). Elles sont

---

<sup>17</sup> APOGEE est une application conçue pour la gestion de la scolarité. <http://www.amue.fr>

conformes aux standards en vigueur, norme Dublin Core, ETDMS<sup>18</sup> ou spécifiques au système de Métadonnées renseignées dans cyberthèses. OGET assure l'export vers les bases de diffusion dans tous les formats possibles HTML, PDF, etc. Le système enregistre l'archivage des supports fournis par le doctorant, et le téléchargement de tous les documents intermédiaires (sources, impressions, illustrations, police) sur un serveur dédié et selon une organisation définie.

Les applications qui viennent d'être présentées sont autant d'outils potentiels pour la mise en œuvre du dispositif de mise en ligne de l'université Paris III. Il convient à présent de les analyser à la lumière des objectifs et des spécificités locales. C'est l'objet de la deuxième partie de cette étude.

---

<sup>18</sup> ETD-MS Metadata Standard for Electronic Theses and Dissertations, ensemble de métadonnées descriptives pour les thèses maintenu par le NDTLD.

## **Partie 2 DU DEPOT A LA DIFFUSION : NOUVELLES PROBLEMATIQUES BIBLIOTHECONOMIQUES**

### **1. Le projet de mise en ligne des thèses à Paris III**

Avec la mise en ligne, la thèse devient donc un document électronique. Toutes les étapes de son « cycle de vie » peuvent être dématérialisées. Toutefois, elle reste un document spécifique. Travail universitaire, classé dans la catégorie de la littérature grise sanctionné par l'obtention du grade de Docteur, elle est aussi une œuvre de l'esprit soumise aux droits d'auteur pour elle-même et pour les documents qu'elle cite. Ces trois aspects : informatiques, administratifs, et juridiques conditionnent le circuit de traitement qui doit nécessairement être réorganisé.

#### **1.1. Repenser le circuit des thèses**

##### **1.1.1. Le circuit actuel**

Au préalable, il est nécessaire de faire un rapide état des lieux du circuit actuel. Au cours de leur cursus, les doctorants doivent passer par différents services :

- le service de la scolarité situé sur le site de Censier, 11 rue de Santeuil enregistre chaque année l'inscription administrative.
- les secrétariats de chaque UFR, sur le même site, assurent les formalités d'ordre pédagogique.
- le service des doctorats, situé au 17 rue de la Sorbonne gère les aspects administratifs liés à la soutenance : organisation, constitution du jury et collecte de tous les formulaires associés (attestation de soutenance et de diplôme par exemple).
- le service des thèses du SCD, localisé à la bibliothèque universitaire, sur le site de Censier, a la charge des opérations liées au traitement du document.

Six semaines avant la soutenance, le doctorant dépose à la bibliothèque deux exemplaires papier de sa thèse, accompagnés d'une disquette comportant les résumés en français et en anglais<sup>19</sup>. Il effectue lui-même la reproduction et l'envoi des exemplaires à chacun des membres de son jury. A l'issue de la présentation, le jury signe le formulaire d'enregistrement de thèse soutenue, autorise ou non la communication, et dresse, s'il y a lieu, un avis sur les corrections à apporter au document. Le SCD le signale dans le Système universitaire de documentation (SUDOC) dès réception de l'avis du jury et du formulaire d'enregistrement de thèse soutenue. Le doctorant dispose d'un délai de trois mois pour effectuer les corrections demandées et déposer les exemplaires définitifs à la bibliothèque.

Le premier exemplaire sert à la communication ; il est équipé d'une étiquette portant la cote, d'un code à barres et d'un antivol puis stocké en magasin. Le second est envoyé à l'ANRT de Lille pour la conservation et la diffusion sous forme de microfiches. Ces microformes sont envoyées dans les bibliothèques de disciplines proches ou dans celles qui en font la demande.

### 1.1.2. Les objectifs à atteindre

Le projet concerne la production courante, soit une moyenne de 160 thèses par an. La rétroconversion des thèses déjà entrées dans le fonds n'est pas prévue pour l'instant.

Quels sont les objectifs généraux ?

- améliorer la diffusion en autorisant une consultation simultanée par les différents utilisateurs potentiels, sur place et à distance
- multiplier le potentiel d'exploitation en permettant une recherche fine en plein texte
- améliorer la conservation
- valoriser la thèse en intégrant d'autres médias (images animées, séquences sonores...)

---

<sup>19</sup> Cela permet une saisie plus rapide des résumés dans le champ 330 de la notice Unimarc de signalement du SUDOC.

Les effets induits :

- augmentation de la consultation du fonds
- diminution des communications depuis le magasin
- diminution de la circulation par le PEB
- augmentation de la qualité de la recherche
- amélioration de la qualité des citations et de la traçabilité des documents

## **1.2. Les formats en question**

Aujourd'hui les doctorants rédigent leur mémoire à l'aide d'un logiciel de traitement de texte. La version éditée sur papier est produite à partir de cette forme native. On peut donc disposer facilement des fichiers numériques sources. Cela constitue un réel avantage par rapport à d'autres projets pour lesquels la numérisation est un préalable nécessaire.

En mettant les thèses en ligne, l'université s'engage dans un nouveau métier. Elle devient responsable de la diffusion électronique du document et de sa pérennité. Elle se trouve ainsi face à des choix radicalement nouveaux. Celui du format informatique est l'un d'entre eux. Dans ce choix, se joue à la fois la diffusion, les usages et la conservation des thèses. Néanmoins, tous les dispositifs informatiques ne répondent pas de la même manière aux objectifs qui viennent d'être exposés. Les outils actuellement en phase de production utilisent principalement deux formats : le format PDF et le format XML.

### **1.2.1. Le format PDF**

Le format PDF *Portable Document Format* a été développé par la société Adobe systems Inc à partir du format d'impression PostScript. Il est très utilisé pour la diffusion sur Internet car il conserve fidèlement les caractéristiques du support papier. De nombreux projets de diffusion électronique des thèses l'ont choisi pour la production, la diffusion et l'archivage. Pourtant, depuis sa création en 1993, il a subi six modifications. On peut donc s'interroger sur sa pérennité. La dernière d'entre elles toutefois, intégrée dans le format PDF(A) permet d'archiver les métadonnées. C'est un résultat encourageant en terme de conservation.

### 1.2.2. Le langage XML

Le langage XML *Extensible Markup Language* a été développé en 1998 par le World Wide Web Consortium, W3C à partir du langage SGML (*Structure Generalized Markup Language*). Il s'agit d'un langage de balisage utilisé pour représenter les données textuelles et les structures logiques des documents. Il a été créé pour être facilement utilisé sur le web et supporté par une grande variété d'applications. Il permet d'échanger les données.

Un fichier XML complet se compose d'une description de la structure (DTD ; maintenant appelée schéma XML) qui définit les balises de texte utilisées et des données proprement dites. Parmi les DTD existantes, on peut citer : EAD, *Encoded Archive Document* pour les documents d'archives, Docbook pour les documents électroniques et enfin la TEI, Text Encoded Initiative pour les données textuelles.

XML permet de créer des langages de balisage adaptés au type d'information à décrire par exemple : XHTML pour les applications Internet, MathML pour les mathématiques sur le web, SVG pour les graphiques, etc. Dernière caractéristique qu'il faut noter, le langage XML intègre UNICODE et peut donc restituer un document dans sa langue d'origine.

PDF	XML
Format en voie de normalisation	Format public
	Format pivot, structuration du document indépendante de sa représentation ce qui rend la diffusion possible dans un format autre que celui utilisé pour l'archivage (par ex. HTML ou PDF)
Pérennité non avérée	Archivage pérenne
Diffusion large (logiciel de lecture Acrobat reader déchargeable gratuitement sur de nombreuses plateformes informatiques) mais lourd à charger pour l'utilisateur	Portabilité et indépendance vis à vis de la plate-forme.
Permet de restituer et d'imprimer le document à l'identique. Protection possible contre la modification, l'impression (car limitée à la copie d'écran), le copier/coller. Cryptage possible (consultation avec mot de passe) mais non garantie par	Permet de restituer les sons, les graphiques, les images animées...  Pas d'option de protection (la diffusion doit se faire dans un autre format)

Adobe	
Adapté au web statique	Adapté au web dynamique
Recherche plein texte peu précise Reprise possible de la structure hiérarchique du document	Indexation possible. Recherche avancée. Affichage souple (tris, sélections, réorganisations etc)

Tableau comparatif des formats informatiques PDF et XML

### 1.2.3. Le choix de Paris III

XML apparaît clairement comme le format le plus adapté au projet de diffusion électronique des thèses car il permet un archivage pérenne tout en autorisant une diffusion fidèle et souple en PDF ou HTML par exemple. Il combine les avantages des deux formats en terme de diffusion et de conservation. Néanmoins, il introduit une contrainte très forte dans l'ensemble du circuit : celle de sa production. En effet, pour que le document soit correctement converti, l'utilisation d'une feuille de style est incontournable. Dans cette nouvelle configuration, les logiques de l'université et celle de l'auteur sont divergentes. Il importe à l'université que le document soit parfaitement structuré pour permettre la conversion au format XML qui offrira aux lecteurs les meilleures possibilités d'interrogation, et d'exploitation des résultats. En revanche, utiliser une feuille de style définie par le SCD peut apparaître inutile et fastidieux à l'auteur au regard du contenu intellectuel et scientifique de son travail.<sup>20</sup> L'utilisation d'une feuille de style fait partie de la rédaction du texte. On pourrait même dire qu'il participe à la réflexion. Pour la réussite d'un tel projet, il est donc indispensable d'impliquer les doctorants dès le début de leur parcours.

Le choix du format XML comme élément structurant entraîne une modification du mode de production, la mise à disposition de feuilles de style, de nouvelles modalités de dépôt, l'installation d'une chaîne de conversion, la définition de sites de diffusion et de modalités de conservation. Au total, il demande plus de temps et de moyens qu'un projet utilisant le format PDF. Il faudra diffuser cette nouvelle culture au sein de la communauté universitaire (étudiants et personnels) en prévoyant des campagnes d'information, des sessions de formation à l'utilisation

---

<sup>20</sup> Démontrer que l'usage de la feuille de style est un réel atout pour le doctorant en cours d'étude comme mode de structuration de la réflexion et au moment de la rédaction finale, pour l'édition du document sera un des arguments à développer lors des formations.

avancée d'un logiciel de traitement de texte ainsi qu'une sensibilisation aux problèmes de droits d'auteur.

En dépit de ces contraintes, le projet de mise en ligne des thèses de Paris III s'appuiera autant que possible, mais de façon pragmatique sur le format pivot XML.

### **1.3. La nouvelle organisation**

#### **1.3.1. UniTEL : la collaboration des services universitaires**

Le groupe projet des conservateurs chargé de l'étude administrative et juridique préalable a fait trois propositions au comité de pilotage. Rappelons les principes de chacun d'eux. Les trois scénarii prévoient un dépôt mixte, papier et informatique comme c'est le cas dans l'ensemble des universités qui ont déjà mis en œuvre un nouveau circuit des thèses.

- Le Scénario 1, UniTEL, repose sur le circuit actuel mais renforce la participation des unités déjà impliquées et inclue celle d'autres services transversaux pour assurer notamment l'information, la formation et le stockage.

- Le scénario 2, E-Docte, place les Écoles doctorales au cœur du circuit. Chacune d'elle devient maître de l'ensemble du parcours des doctorants, informe, forme, et prodigue une assistance technique.

- Le scénario 3, TEL Q'L propose une mutualisation des moyens entre les établissements qui coopèrent dans le cadre du projet de la Carte Documentaire du Quartier Latin à savoir : les SCD des universités de Paris 1, Paris 2, Paris 3, Paris 4 et les bibliothèques interuniversitaires de la Sorbonne, Sainte-Geneviève, Cujas, Sainte-Barbe.

Le deuxième comité de pilotage a choisi la première proposition. Le projet de mise en ligne des thèses sera donc celui de l'université Paris III dans son ensemble

baptisé UniTEL. Cela implique de nouvelles pratiques de travail pour de nombreux services.

### 1.3.2. Les moyens et les limites

#### 1.3.2.1. *Les outils informatiques*

DigiTool est un système de gestion des documents informatiques. L'application a été acquise par le SCD pour gérer, contrôler et partager les documents électroniques de l'université. Développé par la société Ex Libris, il s'interface avec l'ensemble de la suite logiciel Ex Libris (ALEPH 500 et MetaLib) qui porte déjà le catalogue collectif et le portail. Basé sur une architecture ouverte, Il supporte les standards internationaux Dublin Core, Marc, UNICODE, Open URL, HTML, XML et les protocoles d'échanges OAI-PMH. Il permet en outre, l'administration des utilisateurs après authentification. La base des thèses électroniques sera construite et gérée sous DigiTool. Elle est baptisée localement PRODIGE.

Le système informatique de gestion de la bibliothèque (SIGB), la base des documents électroniques et le portail d'accès dépendent donc du même fournisseur. Cela devrait faciliter l'administration du système général. Le réseau informatique du SCD, réparti sur cinq sites, comporte actuellement trois serveurs sous windows 2000 et deux sous Unix.

#### 1.3.2.2. *Le personnel disponible*

Le service des thèses du SCD est actuellement composé d'une Bibliothécaire adjointe spécialisée, Violaine Dupoux. Le traitement documentaire : catalogage et indexation ; les rendez-vous avec les doctorants occupent la moitié de son temps. Le volume horaire passé au traitement des thèses semble incompressible, c'est en effet l'université Paris III qui possède et conserve l'exemplaire de soutenance et donc procède au catalogage et à l'indexation du document.

Les développements informatiques internes sont assurés par un ingénieur en informatique du CRI, Haïm Pinhede. C'est à lui que sont confiés les développements informatiques complémentaires. Il possède l'expertise de DigiTool.

Enfin la coordination de ce projet fait partie de mon poste elle se fera en collaboration avec Muriel Hoareau, conservateur chargée de l'administration des systèmes informatiques spécialisés de la bibliothèque.

Actuellement, le service est en mesure de faire face au nombre de thèses déposées, mais la mise en ligne ne se limite pas au signalement. Elle introduit dans le circuit du document de nouveaux traitements, comme la conversion des fichiers en format XML. On sait que le temps consacré à cette conversion peut varier de quelques heures à quelques jours en fonction de la qualité du stylage du document et de ses caractéristiques éditoriales (illustrations, matériel d'accompagnement etc...). Le personnel actuel ne peut pas assurer l'ensemble de ces tâches.

### 1.3.3. Les données variables non connues

Il reste un certain nombre de données techniques et financières dont la bibliothèque n'a pas la maîtrise. Elles constituent les limites du projet. La rubrique « recherche » du contrat quadriennal prévoit une somme consacrée aux thèses électroniques. Mais nous ne savons pas encore quelles seront les dispositions finales du contrat. Seuls les résultats des négociations financières sont pour l'instant officiels. La réalisation effective du projet dépend en tout premier lieu de la décision de conseil d'administration.

Enfin, depuis le début de l'année 2005, l'environnement technologique et réglementaire qui conditionne le projet évolue fortement. Les métadonnées de gestion qui constituent la deuxième partie des préconisations de TEF ne sont pas encore publiées. Elles sont annoncées pour la fin de l'année 2005. L'application STAR commandée par le ministère n'existe pas encore (cf. p. 18).

## **2. Le traitement documentaire : gestion du document, signalement, accès, diffusion et conservation**

Les principes de traitement d'un document électronique comportent quelques spécificités.

## 2.1. La modélisation du workflow

Le suivi d'un document dématérialisé dans le circuit de traitement, est une nouvelle problématique. L'ensemble des informations nécessaires à l'identification, le contrôle administratif et la vie de la thèse en général doivent être collectées. L'objectif du workflow est donc de valider chaque étape et de permettre ainsi à l'opérateur de savoir, pour chaque thèse quelles sont les opérations qui ont été effectuées et ce qui reste à faire (dépôt complet, attente de validation, traitement en cours, type de diffusion, archivage, communication de l'URL final à l'auteur et l'École doctorale etc...). Il doit pouvoir couvrir l'ensemble de la chaîne depuis le dépôt jusqu'à la mise en ligne.

Il recouvre donc :

- le dépôt du matériel informatique
- la conversion
- le signalement
- la diffusion
- l'archivage
- le contrôle et les statistiques

Le workflow doit également répondre aux critères suivants.

### 1. rapidité de traitement

La circulaire ministérielle insiste sur la nécessité de rapidité. Il faut donc chercher à obtenir un circuit le plus fluide possible en évitant la double saisie pour obtenir une mise en ligne rapide.

### 2. normalisation des métadonnées

Les métadonnées ou « données sur les données » correspondent à la description d'un document. Elles font actuellement l'objet d'une normalisation. Le groupe CG46/CN357/GE5 de l'association française de normalisation, AFNOR vient de publier une recommandation spécifique, *Les métadonnées des thèses électroniques françaises TEF* qui décrit la syntaxe des métadonnées descriptives. Elle s'imposera naturellement à l'ensemble des projets existants et à venir.

### 3. automatisation maximale du traitement

Certaines métadonnées proviennent directement du document lui-même ou du matériel collecté. Ce sont des métadonnées que l'on pourrait qualifier de métadonnées internes. Elles sont issues de la page de titre pour les données descriptives ; des fichiers informatiques et des logiciels utilisés lors de la création du document pour les métadonnées techniques. D'autres existent dans des applications externes présentes dans la chaîne de traitement. Leur collecte aux différentes étapes du circuit et dans le module de workflow doit se faire autant que possible de façon automatique et sans redondance.

Les informations faisant l'objet de choix intellectuels comme par exemple l'indexation rameau, ou les droits de diffusion échapperont à ce critère et devront nécessairement être renseignées manuellement.

## 2.2. Le signalement

### 2.2.1. Interopérabilité : le protocole OAI

Le projet de Paris III s'inscrit dans le cadre de l'initiative de Budapest pour l'Accès Ouvert<sup>21</sup>. L'OAI-PMH, *Open archives Initiative's Protocol for Metadata Harvesting* ou protocole OAI est un protocole d'échange entre des fournisseurs de données et un fournisseur de services (par exemple une bibliothèque, un centre de documentation, un portail thématique ou local). « Grâce au protocole OAI, une bibliothèque agissant en tant que fournisseur de données a la possibilité d'offrir une visibilité accrue à ses documents, notamment à ses publications électroniques ou à ses fonds spécialisés »<sup>22</sup>. Les questions posées par la machine et les réponses qui lui sont renvoyées sont « encapsulées » dans des balises XML et transportées par le protocole http du World Wide Web.

Rendre une base documentaire compatible avec le protocole OAI implique d'« exposer » les données Dublin Core non qualifié (schéma oai\_dc, publié par l'Open Archive initiative) et, en complément sous tout autre format XML pertinent. Dublin Core a l'avantage d'être interopérable mais il ne suffit pas à tous les besoins de signalement et de description bibliographique. Les préconisations TEF publiées par le groupe AFNOR ont pour but de contourner ce problème.

---

<sup>21</sup> Accessible sur le site <http://www.soros.org/openaccess/fr/read.shtml>

Pour valoriser réellement les thèses électroniques il est important que la base produite soit conforme aux standards définis par l'Open Archive Initiative (donc compatible avec le protocole OAI). Elle devient, de la sorte interrogeable, en même temps que d'autres archives ouvertes comme un seul et même fonds d'archive. En France, le projet Cyberthèses et les thèses en ligne des grandes écoles de Paris, PASTEL fonctionnent comme des entrepôts OAI. Le logiciel DigiTool de la société Ex Libris sur lequel repose le projet de Paris III permet de développer des services OAI.

### 2.2.2. Les catalogues et les bases de données

La description bibliographique des collections est une mission première des bibliothèques. Sa qualité est la condition fondamentale pour une bonne circulation de l'information. La mise en ligne des documents nécessite une réflexion importante sur leur mode de signalement.

Le signalement des thèses doit être un compromis entre la description minimale partagée sur Internet et la description bibliographique détaillée de rigueur dans le catalogue SUDOC (ISBD, AFNOR). Pour cela, la recommandation TEF définit un jeu de métadonnées pour les thèses françaises à partir du vocabulaire Dublin Core. TEF comprend actuellement, dans sa partie descriptive, soixante-quinze éléments.

Chaque thèse ayant fait l'objet d'un dépôt mixte sera signalée au minimum par :

- Une notice bibliographique par support en format unimarc dans le SUDOC et dans MAESTRO<sup>23</sup>. Chaque enregistrement renvoie vers l'autre de manière à signaler la présence des deux supports (liens réciproques saisis dans les champs unimarc 455 et 456). La notice du document électronique comportera un lien vers le texte intégral dans le champ unimarc 856.
- Une notice Dublin Core dans la base de données PRODIGE. La notice donnera accès au document consultable en plein texte. DigiTool offre la possibilité de lui associer des documents complémentaires. On pourrait ainsi imaginer une séquence animée de la soutenance en regard du texte de la thèse. Pour l'instant,

---

<sup>22</sup> Le protocole OAI et ses usages en bibliothèques / François Nawrocki.

la saisie des métadonnées repose sur la norme Dublin Core. Pour un meilleur signalement il faudrait implémenter les éléments de TEF.

Il serait également utile de signaler les thèses dans les bases bibliographiques spécialisées comme par exemple Francis (Base de données bibliographique réalisée par l'INIST-CNRS en sciences humaines ; sciences sociales, sciences économiques).

Les nouvelles thèses mises en ligne pourront être signalées sur la page d'accueil du SCD par l'intermédiaire d'un fil RSS spécifique ou dans la rubrique « nouveautés ».

### 2.2.3. Les accès aux bases

Grâce au portail VIRTUOSE, MAESTRO et PRODIGE sont interrogeable à distance et avec d'autres ressources électroniques. Un lien direct sera établi depuis la page « catalogue » de MAESTRO vers la page de recherche de PRODIGE. De la même manière, un lien direct pourra être établi sur la page de présentation des thèses accessible depuis le site web de la bibliothèque. La page pourra inclure un A to Z, par titre, auteur, date ; un formulaire de recherche simple ou avancée par auteur, titre, année, École doctorale, numéro national, mots clés, mots du résumé, tous les mots etc... combinables. Un accès vers cette dernière pourra être aménagé à partir des sites web des écoles doctorales. Actuellement, cinq des sept écoles doctorales disposent de sites web dédiés. Enfin les thèses seront accessibles par le portail de l'ABES.

## 2.3. La diffusion

### 2.3.1. Modalités de la diffusion

La diffusion électronique des thèses doit s'appuyer sur une convention qui garantit le respect des droits d'auteur du doctorant. L'autorisation de diffusion est d'abord et nécessairement accordée par le jury de la thèse, puis elle est confirmée par

---

<sup>23</sup> Les thèses sur support papier sont actuellement cataloguées à la source dans le SUDOC. Les règles de catalogage du SUDOC prévoient un signalement distinct en fonction des supports. Le signalement de chaque document est chargé

l'auteur. Cette autorisation n'a pas de caractère exclusif, le doctorant conserve ses droits d'auteur et garde la possibilité de faire publier tout ou partie de sa thèse auprès d'un éditeur.

L'auteur a la possibilité de choisir les modalités de diffusion électronique :

- En autorisant ou non les copies privées sous réserve du respect du droit d'auteur. (on pourra imprimer tout ou partie de la thèse à partir de la consultation.). S'il n'autorise pas la copie : le document sera « verrouillé » ou dégradé de façon à ce que la copie et l'impression soient possibles.
- En demandant une diffusion restreinte sur l'intranet, s'il désire publier chez un éditeur par exemple. Néanmoins, si la thèse est jugée confidentielle, la diffusion sur intranet sera appliquée automatiquement.

Le doctorant dispose à tout moment du droit de retrait.

Une thèse est à la fois un document scientifique, qui appartient à son auteur, et un document administratif. A ce titre, sa diffusion doit être une décision de l'auteur et du Président de l'université (déléguée au jury). Néanmoins, ne pas diffuser ne signifie pas ne pas communiquer. L'autorisation de diffusion, accordée par le jury, ne porte en fait que sur la diffusion sur Internet. La version électronique sera consultable depuis le domaine de l'université comme l'est la version papier.

### 2.3.2. Le format de diffusion

« Un document codé en XML peut être décliné sur tous les supports possibles : Internet, Intranet, papier. Pour mettre en forme un document XML suivant les supports choisis il suffit de lui attribuer une feuille de styles XSL comprenant les règles à respecter spécifiques à chaque support.»<sup>24</sup>

Les thèses pourront être diffusées et consultées, soit en format HTML, soit en PDF. Le premier est un format léger. Il peut donc être consulté sur des équipements personnels qui ne disposent pas de Modem puissants ou de l'ADSL. Le format PDF, format de diffusion le plus courant, respecte la mise en forme du

---

quotidiennement dans le catalogue collectif du SCD.

<sup>24</sup> XML, pour une appropriation du langage par les professionnels de l'information. Documentaliste-Sciences de l'information. 2003, vol. 40, n° 6.

document. Une question se pose alors. Quel fichier proposer ? Celui déposé par le doctorant ou celui issu de la conversion ? En effet, le fichier issu de la conversion a subi des transformations. Il n'est plus véritablement identique au document de soutenance. L'université Lyon 2 propose le choix entre les deux.

## 2.4. La conservation

Le circuit du document s'achève avec les mesures de conservation. Celle-ci justifie à elle seule la longue et complexe conversion en format XML. En effet, en séparant le fonds de la forme du document, le contenu de sa présentation XML aide à préserver l'intégrité des données. XML véhicule à la fois les données et les moyens qui permettent de les déchiffrer. Dans la chaîne locale, il serait bon de stocker les données au CINES et sur un serveur de l'université. Le Centre de ressource informatique de l'université s'est engagé à assurer ce stockage local. L'application de dépôt STAR assurera semble-t-il l'export des fichiers vers un serveur du CINES pour l'archivage national.

## 3. Les outils du projet

Le système de mise en ligne des thèses de Paris III comportera un dispositif de dépôt, de conversion et de workflow. Reprenons les principales caractéristiques des outils conçus ou utilisés dans les expériences françaises étudiées dans la première partie de ce travail :

SYSTEME	Organisation et structuration	Problématique et fonctions
CYBERDOCS Université Lumière Lyon2	Open Source DTD TEI Lite	Publication électronique de documents (thèses essentiellement) Conversion XML Diffusion et indexation via la plate-forme SDX
SPARTE ABES	Système centralisé DTD TEI Lite	Conversion XML seulement workflow intégré
CASTORE École des Mines de	Open source DTD Docbook puis TEI	Publication électronique de documents

Nantes	Lite	Conversion XML Diffusion et indexation Workflow
INJAC ESUP-Portail (Valenciennes, Nancy 2, Nancy 1, Rennes1 et Toulouse)	DTD Docbook	Publication électronique de documents Conversion XML Diffusion et indexation Workflow
OGET	Open source	Dépôt et workflow

Tableau synthétique des solutions existantes

### 3.1. Le choix d'un outil de conversion

SPARTE et Cyberdocs sont les deux chaînes les mieux documentées pour l'instant ; ce sont aussi les plus anciennes. Leur fonctionnement faisait l'objet de l'étude de Six et dix déjà évoquée. Bien que SPARTE soit toujours officiellement présentée comme une solution technique pour la conversion, on peut s'interroger sur son avenir à long terme. En effet, la structuration XML n'est plus expressément recommandée par le ministère. La correction des bugs de cette chaîne n'est pas actuellement la priorité de l'ABES qui se concentre sur le catalogage multi-écritures dans le SUDOC, la définition des métadonnées, le portail et la conception de l'application de dépôt. Sparte n'est donc probablement plus une solution envisageable.

La plate-forme utilisée dans le projet Castor, bien que récente, paraît prometteuse. Les témoignages des utilisateurs montrent une conversion assez souple, et moins soumise à la qualité du stylage du document. Mais on peut se demander quelles seront les conséquences de l'intégration de la TEI Lite sur ce point. De toutes les expériences présentées, c'est donc la plate-forme Cyberdocs qui présente le modèle le plus abouti. Elle est certes Open Source mais elle est aussi modulable et adaptable aux besoins internes. Nous pourrions imaginer d'utiliser seulement les modules de gestion et de production en amont de la diffusion dans Digitool. Aujourd'hui, cyberdocs est la seule chaîne française déployée à l'échelle internationale. Elle est en effet expérimentée par des bibliothèques africaines et déjà utilisée en Amérique latine. Le réseau de compétences et d'expériences qui s'est construit autour d'elle promet d'être riche d'enseignement et lui confère un potentiel d'évolution important.

On testera donc la chaîne de production Cyberdocs. La plate-forme est disponible au téléchargement sur le site collaboratif du CRU, Comité Réseau des Universités : <http://sourcesup.cru.fr/cybertheses/fr/>.

Depuis mai 2005, Violaine Dupoux a collecté des fichiers source. Pour mener les tests, nous disposons déjà de 15 thèses en format word. Il faudra néanmoins prévoir le temps nécessaire au stylage. L'objectif de cette phase est double : choisir l'outil de conversion et se familiariser avec le stylage. Le didacticiel mis en ligne sur le site de Cyberdocs : le *vade mecum* des utilisateurs de la chaîne qui décrit toutes les étapes de traitement et notamment la préparation du document de travail et le stylage est un point de départ.

<http://sophia.univ-lyon2.fr/CyberTheses/preparation.html>

Le site du CRU, comité réseau des universités héberge également deux listes de diffusion : cyberthèses-tech et cyberthèses-users au travers desquelles les utilisateurs de la chaîne échangent leurs expériences. Elles permettent de répondre aux questions relatives à l'installation, le fonctionnement et l'utilisation de cyberdocs.

A ce stade de notre étude, la chaîne de conversion Cyberdocs semble être l'outil de conversion le plus adapté au projet de Paris III. Il doit être complété par un système de workflow comme OGET.

### 3.2. Le signalement, la diffusion et la conservation

Pour assurer les fonctions bibliothéconomiques plus traditionnelles que sont le signalement, la diffusion et la conservation, le projet dispose des outils suivants.

Objectifs	Critère	Application
Signalement	Conformité au format	SUDOC (WinIBW)
	Unimarc	MAESTRO (Aleph 500)
Accès et diffusion	Interopérabilité	Portail SUDOC
	Conformité au protocole	VIRTUOSE (MetaLib)
	OAI	PRODIGE (Digitool)

	Implémentation de TEF	?
Conservation	Pérennité	Serveur du CINES Serveur local du CRI

Tableau synthétique des objectifs et critères auxquels les applications doivent répondre

Il reste néanmoins des questions à résoudre car le lien entre la partie de la chaîne qui assure la production, et celle chargée des fonctions bibliothéconomiques (signalement, accès, diffusion et conservation) devrait être en principe assuré par l'application STAR.

### **3.3. Les questions en suspens: conséquences de la mise en production de STAR**

Le système mis en place dans le projet de Lyon 2, depuis le dépôt dans OGET en passant par la conversion, l'indexation puis la diffusion semble assez intégré. Mais la question de l'introduction de STAR dans ce processus se pose. L'Agence Bibliographique de l'Enseignement Supérieur a été mandatée par le Ministère chargé de l'Enseignement supérieur pour développer une application de dépôt, diffusion et archivage STAR. On ne sait que peu de choses sur ses caractéristiques techniques, mais son utilisation attribuera une valeur officielle à la thèse déposée, convertie ou non au format XML.

Le principe de travail repose sur une saisie collaborative, validée par un administrateur respectant les recommandations de TEF. Une des fonctions de cet administrateur consistera à exporter des métadonnées vers le SUDOC pour le catalogage, et vers les bases de diffusion. Les thèses pourront être déposées et diffusées dans différents formats XML, HTML, PDF etc... STAR assurera également l'export des fichiers vers le CINES pour un archivage pérenne. Un identifiant persistant et un résolveur de liens garantiront l'accès permanent à la thèse officielle. Il permettra ainsi aux utilisateurs des thèses de faire des citations. Dernier aspect : il s'agit d'une gestion centralisée et délocalisée. Les serveurs se trouveront à Montpellier au CINES. Ceci présente l'avantage de décharger les établissements du suivi technique mais cela implique aussi une évolution lente,

permettant peu de paramétrage local. L'université aura probablement assez peu la maîtrise du système, sauf celui d'administrer les utilisateurs.

Comme nous l'avons vu précédemment, la gestion du dépôt et de la conversion du document en amont de son dépôt dans STAR nécessite une application de workflow. C'est la fonction principale d'OGET application développée par L'université Lyon 2 dans le cadre du projet Cyberdocs. OGET et STAR ne répondent pas à la même définition du dépôt. Le premier assure un dépôt initial du matériel informatique ; le second assure un dépôt post conversion mais ils présentent des fonctions redondantes comme par exemple l'export vers les dispositifs de signalement, diffusion et conservation. OGET a l'avantage d'être développé en Open source. Il est installé localement. Nous ne savons pas comment OGET et STAR évolueront et comment ils pourront être associés l'un à l'autre.

Quelques axes de réflexion :

Nous devons attendre la mise en production de STAR pour en connaître les possibilités réelles et savoir comment interfaçer les différentes « briques » du dispositif.

Selon toute hypothèse STAR pourra extraire automatiquement les éléments Dublin Core de la thèse elle-même, ainsi que les métadonnées techniques. Mais que dire de la saisie des autres éléments de TEF comme par exemple les données administratives ?

Le travail « collaboratif » est prévu ainsi pour l'instant : chaque acteur (ou service) alimente les métadonnées qu'il produit en fonction de son « métier ». On peut penser en première analyse qu'il est contre productif au regard des pratiques quotidiennes. En effet, le service des doctorats alimente déjà sa propre base de données (4D). Il paraît difficile de lui demander une ressaisie manuelle même partielle de ces métadonnées dans l'application STAR. Dès lors la solution envisagée consisterait à confier cette saisie au SCD avec une procédure automatique. Dans sa version 2, STAR sera capable d'importer les fichiers de thèses et les métadonnées de TEF à partir des systèmes locaux.

Le service des doctorats de l'université travaille actuellement à une migration des données de la base 4D vers APOGEE. A la demande du SCD, le CRI a déjà mis en place une procédure d'importation de la table des étudiants d'APOGEE vers celle du module d'inscription d'ALEPH. Nous disposons donc de ce savoir-faire comme point de départ. Cependant, rien ne sera possible avant que STAR permette d'importer des données depuis des applications externes.

## **Partie 3 La mise en œuvre**

Compte tenu des incertitudes exposées plus haut au sujet des outils et de leur intégration future, il paraît pour l'instant difficile de présenter une chaîne définitive.

### **1. Mise en place de la chaîne de traitement**

Conformément aux principes adoptés par le SCD, cette partie de mon étude reprend pour une grande part les conclusions remises en juin 2005 par le groupe projet de l'ENSSIB. Mon travail avait pour but d'étudier les conditions de mise en œuvre technique du projet. Celui-ci s'insère dans un contexte en forte évolution technique, juridique, et réglementaire. Les nouveautés annoncées pour janvier 2006 invitent à la prudence.

#### **1.1. Le nouveau circuit des thèses**

##### **1.1.1. Du dépôt à la mise en ligne**

**Avant la soutenance**, le doctorant dépose au service des thèses du SCD

- 2 exemplaires papier
- 1 exemplaire électronique sous forme de disquette, mél, Zip, clé USB, cédérom, accès FTP
- l'attestation de conformité entre les documents papier et électronique.

Le SCD procède :

- à l'archivage de la thèse sur le serveur local (système de nommage des dossiers univoque à déterminer par exemple nom du doctorant et numéro national de thèse)
- à la vérification rapide du stylage et délivre une attestation de dépôt qui permet l'organisation de la soutenance

**A l'issue de la soutenance**, le jury

- demande des corrections si nécessaires
- remet l'attestation de soutenance et l'avis sur les délais de diffusion et communication de la thèse.

**Après la soutenance**, le doctorant dépose :

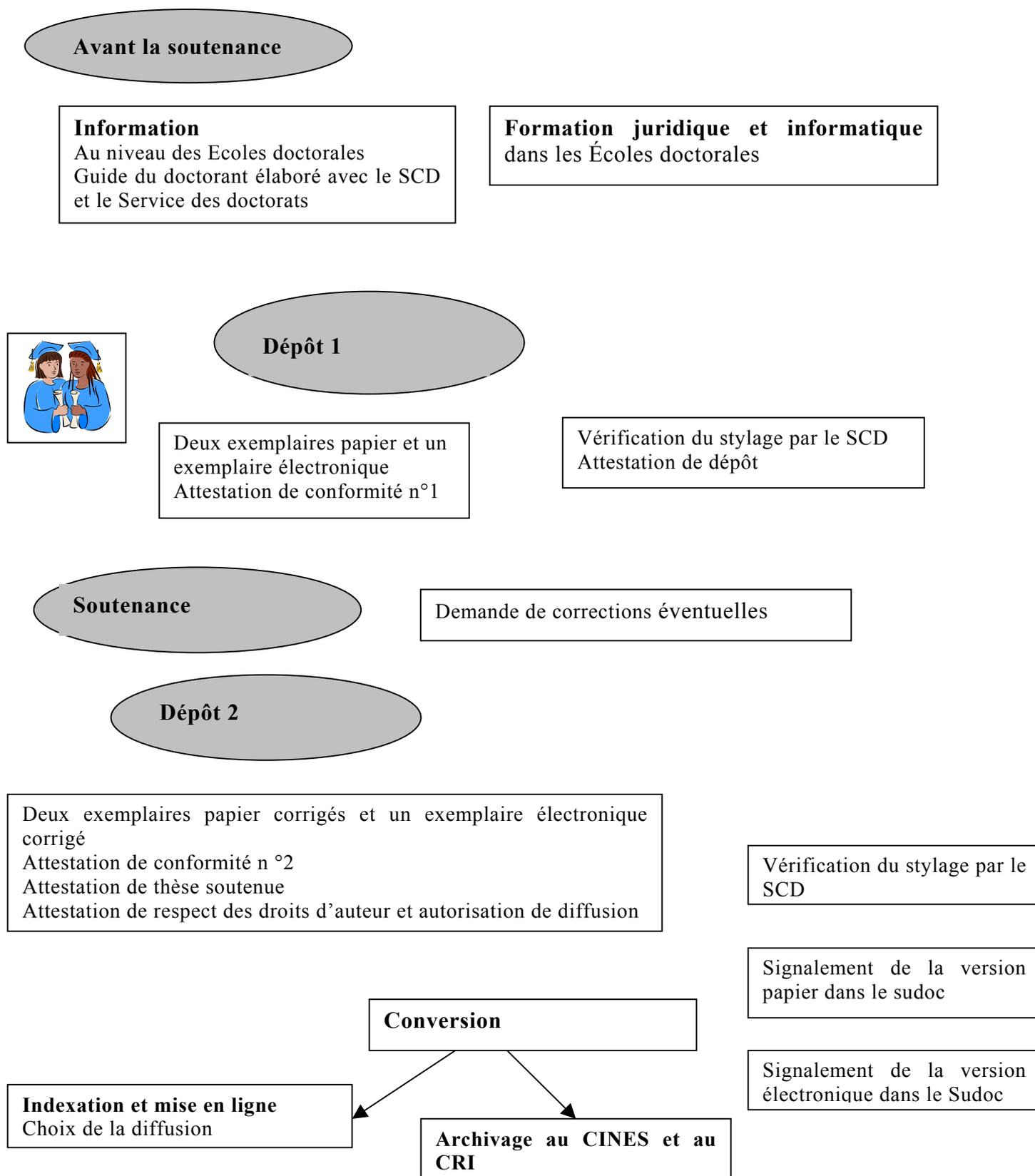
- le contrat de diffusion et l'attestation de respect des droits d'auteur et de citation signés
- 2 exemplaires papier et 1 exemplaire électronique corrigés si besoin accompagnés de la deuxième attestation de conformité.

Le SCD procède :

- à la vérification du stylage du deuxième dépôt
- à l'envoi d'un message au service des doctorats qui atteste du dépôt. Le service des doctorats peut alors envoyer le diplôme au doctorant
- au catalogage de la thèse papier
- à l'évaluation de la durée de traitement/ difficulté et choix du format
- au stylage et conversion
- au catalogage de la version électronique
- au dépôt dans le module de dépôt et archivage de la version électronique

Le processus est validé par le responsable. Il procède à la mise en ligne, l'indexation et la diffusion. La fin du circuit est marquée par l'envoi d'un message électronique au doctorant et au service des doctorats pour communiquer l'URL.

## 1.1.2. Schéma général du circuit



## **1.2. Des métiers, des acteurs : répartition des tâches et implications des différents services**

### **1.2.1. Le doctorant**

Le doctorant suit les formations nécessaires et compose sa thèse en respectant la feuille de style et les instructions. Il procède au dépôt de sa thèse au SCD dans les deux versions accompagné des formulaires demandés. Il effectue les corrections nécessaires et communique au SCD l'attestation de soutenance et l'autorisation de mise en ligne. Il envoie les exemplaires de soutenance aux membres de son jury.

### **1.2.2. Les écoles doctorales**

Les écoles doctorales diffusent les informations nécessaires et orientent les doctorants.

### **1.2.3. Le service des doctorats**

Le service des doctorats diffuse le guide du doctorant, rédigé avec le SCD, et la charte des doctorats. Il effectue ses tâches habituelles : saisie des informations administratives à la thèse, organisation de la soutenance, délivrance du diplôme sur présentation de l'attestation de dépôt.

### **1.2.4. Le jury**

Le jury signe l'attestation de soutenance, demande des corrections si nécessaire. Il décide de la confidentialité et donne un avis sur la diffusion.

### **1.2.5. Le service commun de la documentation**

Il élabore le guide du doctorant en collaboration avec le service des doctorats, forme les formateurs du CAVI, il convertit la thèse, la dépose dans le module de dépôt, indexe et met en ligne, signale les deux versions, archive les différentes versions, gère le site des thèses. Il assure une hotline conjointement avec les intervenants du CAVI. Il coordonne le projet.

### 1.2.6. Le centre audiovisuel et informatique (CAVI)

Il forme les doctorants à l'utilisation de la feuille de style par le biais de ses moniteurs ou intervenants recrutés dans le cadre du C2I niveau 1. Il assure une assistance à la rédaction.

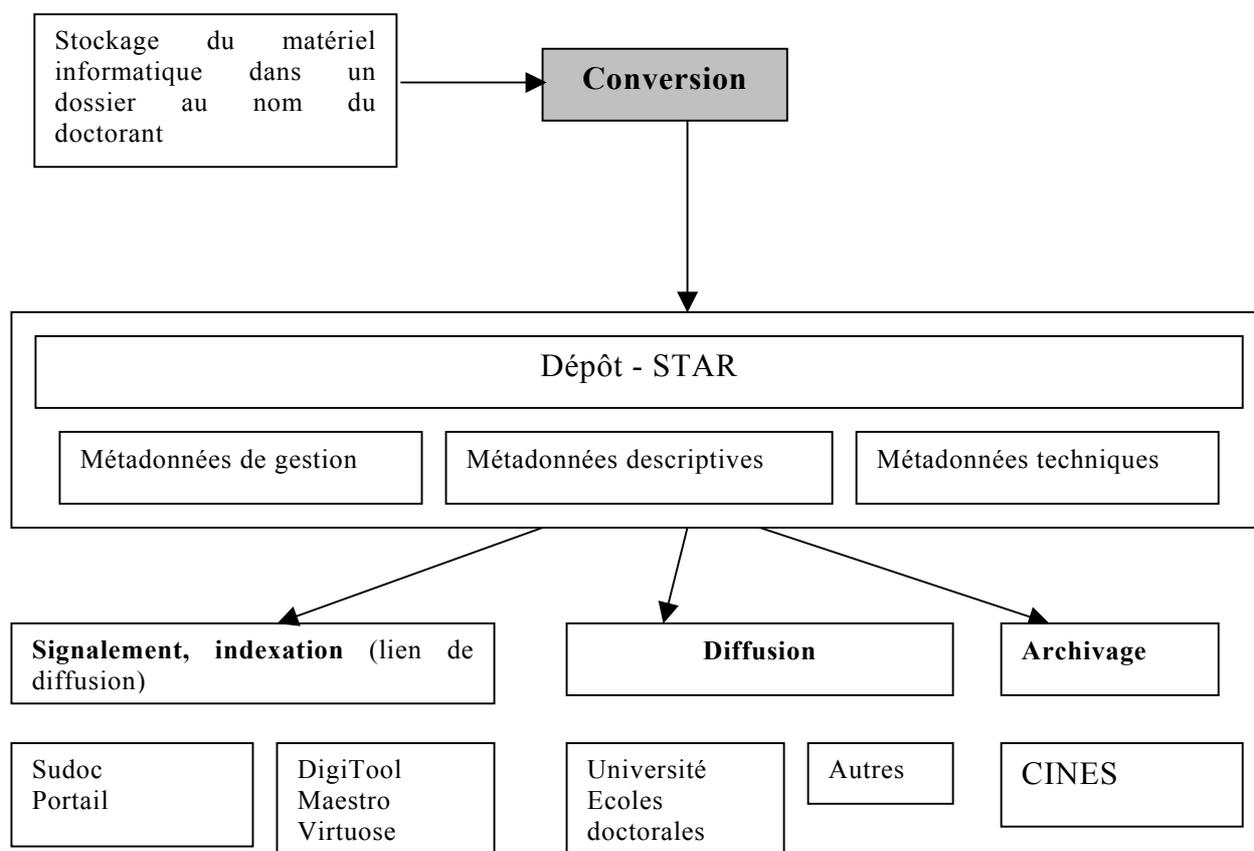
### 1.2.7. Le centre de ressources informatiques (CRI)

Il installe la plate-forme de conversion et la chaîne de traitement, assure l'archivage local.

### 1.2.8. Autres intervenants extérieurs

Un intervenant extérieur recruté par le service juridique de l'université est chargé d'assurer une conférence sur les aspects juridiques de la diffusion électronique des thèses.

### 1.2.9. Articulation des applications : une hypothèse basse sans application de workflow



## 2. Réalisation du projet : programmation

### 2.1. Organisation des tâches au sein du SCD et fiches de postes

#### 2.1.1. Le chef de projet

La coordination de ce projet est incluse dans mes missions. Elle occupera 1/3 ETP régime normal mais un ½ temps au moins, pendant les deux premières phases si l'on considère la rédaction des documents de communication et d'accompagnement le dépôt, la mise en place et les tests de la plate-forme, en collaboration avec le CRI.

Coût : 10160 EUR/an à terme (charges sociales et primes incluses ; augmentées de 35% de charges patronales).

Fiche de poste : (Cadre A conservateur ou bibliothécaire)

- Mise en place, gestion et contrôle de la chaîne de traitement en relation avec le CRI.
- Participation à la rédaction des formulaires, du guide du doctorant en collaboration avec le service des doctorats. Définition du contenu du site des thèses, suivi et mise à jour.
- Participation à la conception des formations des doctorants en lien avec le conservateur chargé des formations et le CAVI. Formation des formateurs
- Encadrement de l'équipe projet des thèses électroniques
- Communication et promotion du projet.
- Veille technologique de la chaîne de traitement (conversion et archivage), suivi des normes et réglementations

Compétences requises : connaissance des serveurs, administration de DigiTool, connaissance de XML et des DTD, connaissance du stylage et de la conversion.

### 2.1.2. Le Technicien

Violaine Dupoux, déjà responsable du service des thèses et du circuit classique assurera le dépôt et la mise en ligne des thèses électroniques. Elle sera également associée à l'installation et aux tests de la plate-forme de conversion. En prenant comme base de calcul un traitement de 200 thèses/an ; 1h30 de traitement par thèse (hors stylage et conversion) ainsi que 75 h/an pour le traitement administratif. Le circuit des thèses électronique représentera une charge de travail supplémentaire de d'1/4 d'ETP.

Coût : 5970 EUR/an

Fiche de poste : (catégorie B ; BAS)

Sous la responsabilité du chef de projet, ses missions seront :

- Participation à la mise en place de la plate-forme, de traitement, conversion, diffusion.
- Réception des doctorants et collecte des documents ; dépôt électronique, signalement, diffusion
- Encadrement du styleur et support

Compétences requises : maîtrise de DigiTool, et connaissance de Dublin Core, Stylage et conversion, catalogage

### 2.1.3. Le styleur

Compte tenu du temps nécessaire au restylage et à la conversion des thèses, il est nécessaire de prévoir le recrutement d'un personnel supplémentaire à temps complet à partir de l'année universitaire 2006-2007. L'usage dans les universités déjà engagées dans ce type de projet est en général de recruter des moniteurs étudiants ou des vacataires de catégorie C. Le temps de traitement de moyen d'une thèse varie entre une demi-journée et une journée selon la difficulté et la fréquence du dialogue avec les étudiants.

Coût : 22100 EUR/an

Fiche de poste : (catégorie C)

Sous la responsabilité du technicien, il aura pour mission :

- Stylage et conversion des documents en XML ou PDF

Compétences requises : maîtrise des outils bureautiques et des logiciels de traitement d'images ainsi que des formats de compression. Maîtrise du logiciel de conversion.

Remarque : pour l'évaluation générale des coût humains et financier du projet voir le mémoire du groupe ENSSIB.<sup>25</sup>

## 2.2. Le matériel

Il est difficile de faire un état du matériel compte tenu des informations techniques manquantes (intégration définitive des différents modules composant le système et application STAR).

On trouve sur le site Cyberthèses/Cyberdocs une description des ressources nécessaires à l'installation de la plate-forme<sup>26</sup>. Il faut « un espace disque d'environ 200 Mo pour une installation complète, incluant OpenOffice.org, PHP, SDX, et java. » précise cette documentation. Or, il est probable que nous n'ayons pas à installer le module de publication qui sera prise en charge par DigiTool. L'espace nécessaire en cours de conversion est évalué à dix fois la taille du document original. Mais la place occupée par le document final peut être plus faible selon les options choisies. Pour évaluer la mémoire vive du serveur, il faut compter 50 Mo par document augmenté de 1,25 Mo par page.

Il faudra certainement prévoir un serveur pour la conversion et la validation des documents 1 pour Digtol et 1 pour la conservation au CRI.

La chaîne sera installée sur 3 postes au SCD et un au CRI. Ces postes clients ont besoin d'un navigateur Web récent supportant les normes HTML, CSS et javascript.

---

<sup>25</sup> Anne Claire Bernaudin, Camille Déguez, Lucile Dupuich [et al]. Projet de mise en place d'un circuit de traitement et de diffusion des thèses électroniques au Service commun de la documentation de Paris III : aspect administratif et juridique.

<sup>26</sup> <http://sourcessup.cru.fr/cyberthèses/fr/installation/materiel.html>

Les logiciels :

La plate-forme est composée de logiciels libre mais le processus nécessite l'utilisation de logiciels complémentaires tels que Microsoft Word pour le re-stylage, la version complète du logiciel Adobe Acrobat pour la production du PDF.

### **3. Le calendrier et le suivi**

#### **3.1. Le «phasage»**

Les réalités techniques et les témoignages recueillis au cours de mon étude, imposent néanmoins quelques aménagements au calendrier proposé par le groupe projet de l'ENSSIB. La mise en route opérationnelle pourra bien avoir lieu en trois étapes mais la programmation est un peu décalée. Celle-ci doit tenir compte de deux paramètres : les réalités techniques, d'une part et le calendrier réel des dépôts, d'autre part. En effet, 2/3 des dépôts ont lieu pendant le dernier semestre, de fin septembre à mi novembre. Le reste des thèses sont déposées au printemps pendant les mois d'avril, mai et juin.

- La phase de lancement peut débuter en janvier 2006. Au dépôt volontaire très aléatoire, on préférera une école pilote : l'École doctorale 268 Langues et langues (environ 20 thèses par an). Madame Vaissière, la directrice semble en effet favorable à la mise en ligne des thèses et conseille déjà ce mode de diffusion à ses étudiants. C'est à cette date que commenceront les formations à l'utilisation de la feuille de style et la sensibilisation aux droits d'auteur. Cette phase de mise en route pourra être préparée durant le dernier trimestre 2005 par une campagne de communication (novembre). L'installation et les tests de la chaîne de conversion pourront débuter dès le mois d'octobre 2005 avec le matériel informatique dont nous disposons déjà. La conversion des thèses de l'école pilote commencera en avril 2006.
- A partir de la rentrée 2006-2007 les formations devront concerner l'ensemble des étudiants de troisième année. On soumettra le vote du dépôt

électronique des thèses au conseil d'administration après la phase de validation. Il devrait pouvoir prendre effet en janvier 2007.

- La troisième phase de « rythme définitif » peut être mise en place à la rentrée 2007-2008. Les formations concerneront les doctorants de première et de troisième année.

### **3.2. Le calendrier**

**Année universitaire 2005-2006**

sept.	oct.	nov.	déc.	janv.	févr.	mars	avr.	mai	juin	juil.	août	
				Phase de lancement - Ecole pilote								
		Réunion avec l'Ecole pilote										
	Installation et test de la plate-forme											
		Formation du personnel										
		Préparation des formations										
				Formation des doctorants								
					Formation des doctorants				Formation des doctorants			
						1ers dépôts						
									Evaluation			
									Information des écoles doctorales			
									Recrutement du styleur			
											Fermeture	

**Année universitaire 2006-2007**

sept.	oct.	nov.	déc.	janv.	févr.	mars.	avr.	mai	juin	juil.	août
Phase de déploiement											
	Formation du styleur										
	Formation des 3e année			Formation des 3e année			Formation des 3e année				
		Evaluation					Evaluation				
		Présentation aux conseils									
				Obligation de dépôt							
											Fermeture

### **3.3. Evaluation et suivi**

L'évaluation consiste à recueillir des informations pour les analyser et les interpréter dans le but de prendre une décision et éventuellement d'agir sur la situation observée.

On peut distinguer deux objectifs principaux : valider la méthode et le calendrier du projet, d'une part, et contrôler son fonctionnement régulier, d'autre part.

La phase dite de lancement sera évaluée en juin 2006, puis fin novembre. Le comité de pilotage pourrait être réuni à cette occasion pour valider ces résultats, avant de les présenter en conseil scientifique et de solliciter un vote du CA pour une extension du projet. Les résultats des deux évaluations donneront éventuellement à des ajustements des objectifs et des moyens mis en œuvre (prolongation de la période de test par exemple). On choisira des indicateurs simples :

- Le nombre de thèses correctement stylées par les étudiants
- Le nombre de thèses effectivement converties

Ces deux indicateurs seront rapportés au nombre de thèses attendues pour l'école doctorale pilote.

A ce stade, évaluer le temps de traitement d'une thèse ne paraît pas très significatif. Cette phase de lancement accompagne en effet, la mise en production de la chaîne de conversion. Le temps d'appropriation de celle-ci va nécessairement majorer cette durée.

Le fonctionnement régulier sera évalué tous les six mois à l'aide des indicateurs suivants :

- Le rapport entre le nombre de thèses autorisées à la diffusion par les doctorants et le nombre de thèses soutenues dans le but d'évaluer l'efficacité la sensibilisation des étudiants.
- Le rapport entre le nombre de doctorants formés à la feuille de style et le nombre de thèses diffusées.

- Le temps moyen de traitement d'une thèse, d'un bout à l'autre du circuit pour évaluer son efficacité, l'adéquation des moyens mis en œuvre et l'efficacité de la formation des doctorants à la feuille de style notamment.
- Le nombre de thèses stylées correctement afin d'évaluer le dispositif de formation
- Le rapport entre les thèses déposées et celles converties et diffusées dans le but de dénombrer le nombre de thèses non convertibles.
- Le nombre de thèses nécessitant un double dépôt de manière à évaluer l'impact de la diffusion électronique des thèses sur la qualité des thèses.

On utilisera enfin :

- Les statistiques de consultation des thèses en lignes pour mesurer la pertinence des objectifs.
- Les statistiques de consultation des supports de formation juridique et informatique sur le site de l'université pour mesurer la pertinence de la formation à distance.

Un rapport synthétique pourra être établi en fin d'année à partir de ces résultats.

### **3.4. Perspectives et avenir**

#### **3.4.1. Au sein de l'université :**

A l'échelle locale, comme le propose le groupe projet de l'ENSSIB, il pourrait être intéressant d'étudier les modalités d'intégration de la problématique de mise en ligne des thèses, dans le dispositif du C2i (Certificat informatique et Internet niveau 2), qui a été programmé par le Ministère pour 2006. En imaginant une spécialisation par discipline, la formation pourrait être déléguée aux Écoles doctorales, et non centralisée.

Avec la généralisation de la diffusion électronique des thèses dans les universités, se posera inévitablement la question du traitement rétrospectif. Il faudra alors mettre en place un programme de numérisation et peut-être recourir à un prestataire extérieur. C'est un gros projet pour un SCD et cela requiert des moyens importants. Pour l'instant, les instructions du Ministère ne mentionnent pas la

reprise d'antériorité. On pourrait imaginer un dispositif encadré par des procédures de marchés comme c'est le cas pour la rétroconversion des catalogues.

Enfin, la collection des thèses électroniques dans DigiTool constitue la première « brique » de la bibliothèque électronique de Paris 3. Le projet servira de base d'expériences et de compétences pour l'utilisation de cet outil, qui a été acquis pour gérer l'ensemble des documents électroniques de l'université, qu'ils soient natifs comme des cours ou des images animées ou bien issus de programmes de numérisation comme des collections d'images. Il s'agira alors, comme DigiTool le permet d'une gestion en collaboration avec les différents services de l'université.

### 3.4.2. Vers un pôle universitaire

A plus long terme, lorsque l'équipe aura acquis une expérience suffisante, le projet pourra être étendu à d'autres établissements parisiens partenaires de manière à mutualiser les moyens et les compétences. Cela permettrait de réaliser des économies d'échelle en fonction des spécialités, comme par exemple un service juridique ressource. Dès lors deux types de partenariats sont possibles :

- le premier s'inscrit dans la prolongation du travail engagé pour la mise en place de la carte documentaire du Quartier Latin.
- Le second s'inscrit dans le cadre de la convention « Alliance Paris Universitas » signée au printemps 2005 par l'université Paris III et l'Écoles des Hautes études en sciences sociales, l'Écoles Normale Supérieure, de la rue d'Ulm, l'université de Paris Dauphine, l'université Pierre et Marie Curie. « Cette convention a pour objectif de définir les modalités de mise œuvre des actions de coopération entre ces cinq grands établissements » affirme le communiqué, la diffusion électronique des thèses pourrait être l'un d'entre eux.

## **Conclusion**

Les projets menés dans d'autres établissements nous fournissent les outils nécessaires à la conversion et au traitement du document. Cependant, les nouveautés annoncées dans la dernière circulaire du Ministère de l'Education nationale et les adaptations techniques qui en résultent invitent à la prudence. Nous ne pouvons donc, pour l'instant, donner de conclusions définitives quant à la chaîne technique de production. Toutefois, cela ne nous condamne pas à l'immobilisme. La conduite de projet est faite de constants réajustements.

Le corpus des thèses sera la première pierre d'une bibliothèque électronique qui rassemblera les documents existants dans les différentes composantes de l'université Paris III. La mise en ligne des thèses mobilisera de nombreux services universitaires. Elle apportera donc au SCD de nouvelles collaborations et une meilleure visibilité de ses actions au sein de la communauté universitaire.

Mais le chemin à parcourir est encore long. La réalisation d'un projet dépend de nombreux facteurs internes et externes qui ne sont pas tous maîtrisés par ses promoteurs. Cette étude illustre parfaitement ce que peut être la théorie à l'épreuve du terrain.

.

# **Bibliographie**

## **La documentation électronique**

**Bourasseau, Céline, Cédric Dumas.** *Enquête sur les pratiques documentaires des enseignants-chercheurs à l'EMNantes.* - [en ligne]. [Dernière consultation le 10-09-2005].

Consultable sur :

<<http://www.emn.fr/x-info/bn/documents/Castor-Etude-E.pdf>>.

**Jaquesson Alain, Rivier Alexis.** *Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux.* Paris : Ed. du Cercle de la librairie, 1999.

**Lupovici Christian.** La chaîne de traitement des documents numériques : caractéristiques et mise en œuvre. *Bulletin des Bibliothèques de France*, 2002, n°1, t. 47.

## **La diffusion électronique des thèses**

**Abdrahamane, Anne.** *Appropriation d'une plate-forme d'édition électronique basée sur XML : Cyberdocs.* Villeurbanne : Enssib, 2003

Rapport de stage : Enssib : Université Claude Bernard Lyon 1 : 2003. [en ligne]. [Dernière consultation le 05-09-2005].

Consultable sur :

<<http://www.enssib.fr/bibliothèque/documents/dessid/rsanne.pdf>>

**Bernaudin, Anne-Claire, Dégez, Camille, Dupuich, Lucile, Feuchot, Nicole, Nguyen, Claire, Nicolas Pinet.** *Projet de mise en place d'un circuit de traitement et de diffusion des thèses électroniques au Service Commun de la Documentation de Paris III : aspect administratif et juridique.* Villeurbanne : Enssib, 2005.

Mémoire de projet : Enssib juin 2005.

**Boulétreau, Viviane ; Université Lumière, Lyon 2, ERAD, Cellule édition électronique.** *Publication et diffusion électronique des thèses : pourquoi ? comment ?*. - [en ligne]. [Dernière consultation le 27-09-2005].

Consultable sur :

<[http://edel.univpoitiers.fr/projet/telecharger/lyon2\\_Formation\\_doc.pdf](http://edel.univpoitiers.fr/projet/telecharger/lyon2_Formation_doc.pdf)>

**Chartron, Ghislaine.** Rapport sur la diffusion électronique des thèses. *Bulletin des bibliothèques de France*, 2001, n°2, t. 47

**Jolly, Claude (réd.).** *Rapport sur la diffusion électronique des thèses*. - [en ligne]. [Dernière consultation le 5-09-2005].

Consultable sur :

<<http://www.sup.adc.education.fr/bib/Acti/These/jolly/entete.htm>>

**Salah, Soraya.** *Mise en ligne en texte intégral des thèses de l'université Paul-Sabatier*. Villeurbanne : Enssib, 2004.

Projet professionnel personnel : Enssib : 2004. [en ligne]. [Dernière consultation le 05-09-2005].

Consultable sur :

<<http://www.enssib.fr/bibliotheque/documents/ppp/pppsalah.pdf>>

**SCP communication.** *Les pratiques des utilisateurs du signalement des thèses et des utilisateurs de thèses : synthèses*. Maisons-Alfort : SCP communication, 1997.

### **Normes et réglementation**

**AFNOR, groupe (CG46/CN357/Groupe 5).** *TEF : métadonnées des thèses françaises*. - [en ligne]. [Dernière consultation le 06-06-2005]

Consultable sur :

<<http://www.abes.fr/abes/documents/tef/index.html>>.

**DCMI Dublin Core Metadata Initiative.** - [en ligne]. [S.l.] : Dublin Core Metadata Initiative, 2003. [Dernière consultation le 19-09-2005].

Consultable sur :

<<http://dublincore.org>>

**Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche.** *Circulaire du 21 septembre 2000 : Diffusion électronique des thèses.* n° 2000-149, BOEN, n°34, 28/09/2000.

**Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche.** *Circulaire du 29 mars 2005 : dépôt, signalement, diffusion et archivage des thèses sous forme électronique.* n° 05-094. -[en ligne]. [Dernière consultation le 13-09-2005].

Consultable sur : <<http://www.sup.adc.education.fr/bib>.>

#### **Formats informatiques et protocoles**

**Nawrocki, François, Ministère de la culture et de la communication, Direction du livre et de la lecture.** *Le protocole OAI et ses usages en bibliothèques.* - [en ligne]. [Dernière consultation le 08-06-2005].

Consultable sur :

< <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm>>

**Quint, Vincent.** *XML, le web et ses documents.* - [en ligne]. [S.l.] : ENSIMAG, 05. [Dernière consultation le 26-09-2005].

Consultable sur :

<<http://wam.inrialpes.fr/courses/ENSIMAG-Jan05/Overview.html>>

*Extensible Markup Language (XML) 1.1, W3C recommendation.* [en ligne]. [S.l.] : W3C, 2004. [Dernière consultation le 26-09-2005].

Consultable sur : <<http://www3.org/TR/2004/REC-xml11-20040204/>.>

**Savourat Lise.** XML, pour une appropriation du langage par les professionnels de l'information-documentation. *Documentaliste-Sciences de l'information*, 2003, vol. 40, n° 6 p. 362-369.

## Documentation pratique

*Castore.* - [en ligne]. [S.1] : École des Mines de Nantes. [Dernière consultation le 26-09-2005].

Consultable sur :

<<http://www.emn.fr/x-info/bn/index.html>>

*Network Digital Library of Theses and dissertations.* – [en ligne]. [S.1] : NDTLD, [s.d.]. [Dernière consultation le 26-09-2005].

Consultable sur :

< <http://www.ndltd.org/index.en.html>>

*Plateforme Cyberdocs : site de développement collaboratif.* [S.1] : Cyberthèses, 2003. [en ligne]. [Dernière consultation le 19-09-2005].

Consultable sur :

<<http://sourcesup.cru.fr/cybertheses/index.html>.>

## ***Table des annexes***

<b>ANNEXE 1 IMPLANTATION DES SITES DE PARIS III .....</b>	<b>72</b>
<b>ANNEXE 2 STATISTIQUES DE CONSULTATION DES THÈSES CONSERVÉES EN MAGASIN .....</b>	<b>73</b>
<b>ANNEXE 3 GLOSSAIRE ET TABLEAU COMPARATIF DES SOLUTIONS TECHNIQUES .....</b>	<b>74</b>
<b>ANNEXE 4 GLOSSAIRE .....</b>	<b>76</b>

## **Annexe 1 Implantation des sites de Paris III**

Les différents sites de l'université



## ***Annexe 2 Statistiques de consultation des thèses conservées en magasin***

	CTLES	Cote TP		Cote TPE		Microfiches	
Année	Nombre de thèses communiquées	Nombre de thèses communiquées.	%	Nombre de thèses communiquées	%.	Nombre de thèses communiquées	%
2001	43	1946	30,0	100	1,5	792	12,0
2002	104	1984	32,0	170	3,0	525	8,5
2003	131	2190	33,0	90	1,1	358	6,0
2004	148	2278	31,0	190	3,0	239	3,0
<b>total</b>	<b>426</b>	<b>8398</b>		<b>550</b>		<b>1914</b>	
<b>Moyenne</b>			<b>31,5</b>		<b>2,2</b>		<b>7,4</b>

Les thèses conservées au CTLES sont des thèses dont la date de soutenance est inférieure ou égale à 1985

## ***Annexe 3 Glossaire et tableau comparatif des solutions techniques***

**CASTORE** : plate-forme de bibliothèque numérique développée par l'École des Mines de Nantes dans le cadre du projet castor. Elle est destinée à la création d'archives de littérature grise.

Type : logiciel public

**CYBERDOCS** : plate-forme de traitement et de diffusion des thèses développée par l'université Lumière Lyon 2 dans le cadre du projet Cyberthèses.

Type : logiciel libre

**DIGITOOL**: application de gestion des documents numériques développée par la société Ex Libris, choisie par le SCD de l'université Paris III pour la constitution et la gestion de la base des thèses et des documents électroniques. Elle est baptisée localement PRODIGE.

Type : logiciel propriétaire

**INJAC** : plate-forme de gestion et de publication de documents développée dans le cadre du projet ESUP-Portail. Elle appartient à la catégorie des produits de CMS (Content Management system).

**OGET** : application de dépôt et de gestion des thèses électroniques développée par l'université Lumière Lyon 2 pour compléter la plate-forme CYBERDOCS.

Type : logiciel public

**SPARTE** : plate-forme de conversion développée par l'Agence Bibliographique de l'Enseignement Supérieur

Type : logiciel propriétaire

**STAR** : application de dépôt, gestion et archivage des thèses commandée par le Ministère de l'Education nationale. Son développement a été confié à l'Agence Bibliographique de l'Enseignement Supérieur.

Type : inconnu pour l'instant

Tableau récapitulatif des outils présentés et de leurs fonctions

Systèmes	Fonctionnalités					
	Conversion	Dépôt	Diffusion	Indexation	Archivage	Workflow
CASTORE	×	×	×	×	×	
CYBERDOCS	×		×	×		
DIGITOOL		×	×	×		
INJAC	×		×	×		×
OGET		×			×	×
SPARTE	×					×
STAR		×	×		×	?

## **Annexe 4 Glossaire**

**Accès ouvert**, *Open Acces* : mouvement international qui vise à permettre à tous un accès à la documentation savante, et notamment aux publications validées par les pairs (*peer reviewed*), en la diffusant gratuitement sur Internet.

**Dublin Core** : ensemble d'éléments destinés à décrire des ressources en réseau. Le Dublin Core simple comprend 15 éléments. Le Dublin Core qualifié est étendu à d'autres termes qui permettent de préciser et de compléter la description. Depuis 2003, le Dublin Core simple est la norme ISO 15 836.

**DTD**, *Document Type definition* : modèle qui permet définir la structure et le vocabulaire de balisage à employer pour un document XML ou SGML particulier. voir aussi: Schéma XML.

**ETD-MS**, *Metadata Standard for Electronic Theses and Dissertations* : ensemble de métadonnées descriptives pour les thèses électroniques maintenu par l'organisme NDTLD. ETD-MS a pour vocation l'échange international des métadonnées de thèses électroniques.

**HTML**, *Hypertext Markup Langage* : langage de balisage dérivé de SGML utilisé principalement pour la création de pages web sur Internet.

**JAVA** : langage de développement, produit par la société Sun et lancé le 23 mai 1995. Écrit par James Gosling, il permet de créer des applications autonomes et de doter les documents html de nouvelles fonctionnalités : animations interactives, applications intégrées, modèles 3D, etc.

**OAI**, *Open Archive initiative* : organisation qui s'est donnée pour but de mettre en place l'interopérabilité des archives. Pour ce faire elle a défini le protocole OAI-PMH, *Open Archive Protocol for Metadata Harvesting*.

**Omnimark** : langage de programmation utilisé pour les tâches de transformation et de traitement SGML et XML.

**PDF** : *Portable document format* : format de fichier créé par la société Adobe et lisible sur une grande variété de plate-formes informatiques.

**Schéma XML** : document XML qui définit les contraintes auxquelles doivent satisfaire d'autres documents XML pour être valides. Ces contraintes concernent les éléments et les attributs XML, les types de valeurs qu'ils peuvent prendre, leur cardinalité et leur espace de noms.

**SGML**, *Structure Generalized Markup Language* : langage normalisé de balisage développé principalement par Charles Goldfarb. C'est aujourd'hui une norme ISO.

**SVG**, *Scalable Vector Graphics* : langage de description d'ensemble graphiques vectoriels basé sur XML.

**TEI**, *Text Encoding Initiative* : ensemble de recommandations destiné à échanger des données électroniques textuelles en sciences humaines. La DTD TEI Lite est une version simplifiée de la TEI dont elle reprend les principaux éléments et attributs.

**XML**, *eXtensible Markup Language* : méta langage dérivé de SGML défini par le consortium W3C.

**XSL** : *eXtensible Stylesheet Language* : langage associé à XML qui permet de décrire des feuilles de style. XSL comprend des sous-ensembles permettant d'effectuer des traitements et des transformations sur les données XML comme par exemple : XSLT, *XSL Transformation*, XSL-FO, *XSL Formatting objects*.